# Characterisation of alternative H2A.X mRNAs during the cell cycle and under DNA damage conditions

A thesis submitted for the degree of

*Doctor of Philosophy in Pathology*

Hilary Term 2019

Supervisor: Nicholas Proudfoot




**Esther Griesbach**

Sir William Dunn School of Pathology

Lincoln College

University of Oxford

# Abstract

Replication-dependent (RD) histone genes employ a unique RNA processing mechanism that generates non-polyadenylated messenger RNA (mRNA), ending in a stem-loop (SL) structure. This is mediated by a U7 snRNP-dependent mechanism that distinguishes this mRNA class from "bulk" poly(A)$^+$ mRNA. However, recent studies have shown that RD histone genes express a subset of polyadenylated mRNA in differentiated cells, as a source for replacement histones. My results now show that the formation of read-through poly(A)$^+$ histone mRNA from RD histone genes is not only a result of the differentiation program, as highly proliferating HeLa cells also express some of these mRNA isoforms. Furthermore, these poly(A)$^+$ histone mRNA isoforms in HeLa cells were also detected in the cytoplasm making them potential translation substrates.

The gene encoding for H2A.X, H2AFX, is the best known histone gene that expresses both SL and poly(A)$^+$ mRNA isoforms, so showing replication-dependent and replication-independent characteristics. My data indicate that the H2A.X poly(A)$^+$ mRNA is mainly generated by H2AFX transcription outside of the S-phase. A minor fraction may be produced in S-phase as a consequence of inefficient U7-dependent processing. Furthermore, H2A.X poly(A)$^+$ mRNA levels are upregulated in G2/M in both HeLa and HCT116 cells.

By specifically depleting either of these two mRNA isoforms, using RNAi and CRISPR-Cas9 gene editing, I show that their requirement for efficient $\gamma$H2A.X foci formation upon DNA damage induction depends on their relative levels. Thus cells that express lower levels of the H2A.X SL mRNA, such as HeLa and RPE-1 cells, mainly depend on the H2A.X poly(A)$^+$ mRNA for *de novo* H2A.X protein production outside of S-phase but possibly also for sufficient H2A.X incorporation into newly synthesised chromatin during S-phase. In contrast, cells that express high levels of the H2A.X SL mRNA, such as HCT116 cells, mainly depend on this mRNA isoform. In these cells, the incorporated H2A.X in the chromatin during S-phase is sufficient for efficient $\gamma$H2A.X signalling following DNA damage induction.

Only vertebrate H2AFX genes have evolved to contain the SL structure. Apart from allowing formation of H2A.X SL mRNA by U7-dependent processing, the function of the SL in the H2A.X poly(A)$^+$ mRNA was unknown. I now show that the SL structure has a role in restricting H2A.X poly(A)$^+$ mRNA accumulation in S-phase, possibly through interaction with the conserved region upstream of the polyadenylation site.

# Acknowledgements

First of all I want to thank you, Nick, for taking me on as your DPhil student number 32. It has been an incredible journey. I've had ups and of course also downs and I'm extremely grateful to you for helping me through those. Thank you for being so generous, for letting me follow my passion and for giving me freedom. Also, thank you for pushing me when I needed it, for listening to my worries and helping me to get back onto my path when I had deviated from it. I don't know what is coming next, but what I know is that it is going to be really difficult, if not impossible, to find a mentor and friend like you. Thank you for making the lab such a great place.

This project wouldn't be the same without the guidance from Bill Marzluff (University of North Carolina at Chapel Hill, USA). Thank you so much for letting me visit your lab. Working beside you was a true inspiration. And thank you for always being there for me when I needed advice, especially also after my time in Chapel Hill. Also, thank you so much for proofreading my thesis. I am deeply grateful for all your help and motivation. I really hope we can keep this collaboration alive in the future.

A special thank you goes to Chris Norbury, thank you for always making time for me when I needed your advice, I am deeply grateful for your help over the years. I also want to thank Elizabeth Bikoff, Elizabeth Robertson and Anna Senft. Thank you for sharing equipment that has saved me many hours of gel loading and of course for your support in many other ways during the last four years.

A huge thank you goes to all the people that have given me technical advice. Thank you Anna-Lisa Lanz, Giulia Masi and the entire Acuto lab for your help with the Licor system. Alan Wainman, thank you for teaching me microscopy and for always being there when I had a problem with the microscope or my image analysis. Thank you Helio Duarte Roque for writing the image analysis script for me that allowed me to analyse my images in an unbiased way. Thank you Lewis Arthurton for your help with other ImageJ scripts and Maryam Qurashi for your help with the statistical analysis of my images. I also want to thank Michal Maj and Line Eriksen, thanks for your help with the flow cytometers and the sorting, I couldn't have done all the flow cytometry without you! A big thank you also goes to Jacob Matson and Jean Cook (both at University of North Carolina at Chapel Hill, USA), for sharing their EdU labelling protocol for flow cytometry. Also, thank you Isabel Wassing and Chris Holmquist (University of North Carolina at Chapel Hill, USA) for sharing your DTB protocols. Thank you Marc-David Ruepp (University of Bern, Switzerland), for the histone promoter-EGFP plasmids and Matthias Gromeier, Elena Dobrikova (both at Duke University Medical Center, North Carolina, USA) and Alfredo Castello (Department of Biochemistry, University of Oxford, UK) for the HeLa Flp-In T-REx cell line. Thank you Nathaniel Moorman and Erik Lenarcic (both at University of North Carolina at Chapel Hill, USA) for your help with the polysome gradient fractionation. Thank you Xanita Saayman and Chris Carnie for showing me how to do PFGE. I'm extremely grateful to Joey Riepsaame for his help with my CRISPR experiments, CRISPR changed this entire project! A huge thanks also goes to Lewis Taylor and Sophia Koepke (Department of Engineering Science, University of

Oxford, UK), without you I wouldn't have written my thesis in LaTex. Additionally, I want to thank everyone of the Proudfeet for endless advice throughout the years, as well as all the wonderful people in the Dunn School that have offered great help every time I needed it.

Next I want to thank all my close friends that helped me along the way. Thank you Hannah Mischo for inspiring and motivating me. You always knew how to make me feel better and you made me see the light where I couldn't see it. Thank you for being there when I needed you the most. And, very importantly, thank you so much for proofreading this thesis even when you didn't really have time to do it. I also want to thank my dear friends Adam Volanakis and Rita Krawczyk, for immediately taking me in and for standing by me every day of this journey - without you the lab and Oxford in general would have been a much less fun place. Rita, without you I wouldn't have been able to analyse my RNAseq data, thank you for guiding me and for being so patient with me, I will never forget that. I also want to thank my very close friends Elisa Ramos, Carlota Recio and Angela Moncada. Thank you for always listening and advising me. I knew I could always count on you. You made me feel more at home here. I also want to thank my dear friend Anna-Lisa Lanz, especially for your support during this last year. Thank you for always being there for me. Thank you Joan Monks, for making my work in the lab possible and for always listening to and helping me through my worries. And of course for the "good-morning-chats" that have made every start into the day brighter. Sue Mei Tan-Wong, my dear deskmate and true friend, thank you for always being there for me and for never saying "no" when I needed your help. I don't know what I would have done without you. I also want to thank you Mariya Lobanovska, thanks for your help with the submission and for all the wonderfull lunch breaks. A warm thank you also goes to my friend Valentina Romeo, thank you for your advice and for always cheering me up.

También quiero dar las gracias a mis mejores amigas Susana Hidalgo, Aina Maria Fiol, Marta Castellano, Anna de la Salud y Sofía Izquierdo. Gracias por siempre estar a mi lado, independientemente de lo lejos que estemos físicamente. Mi vida no sería tan bonita sin vosotras. Gràcies.

Das grösste Dankeschön geht an meine Familie und an Dich, Matthias. Ohne euch hätte ich es nicht bis hierher geschafft. Danke, dass ihr immer für mich da wart, und dass ihr mir durch die schwersten Zeiten geholfen habt und immer an mich geglaubt habt. Lieber Matthias, was hätte ich nur ohne Dich gemacht? Vielen Dank, dass Du mit mir durch Dick und Dünn gegangen bist. Und für Dein Verständnis. Nie war es ein Problem für Dich, wenn ich am Wochenende arbeiten musste, oder wenn ich nicht so lange Urlaub machen konnte wie wir gerne gewollt hätten. Ausserdem, werde ich dir Deine unermüdliche Reiseausdauer und Deine unendliche Grosszügigkeit nie vergessen! *Aaf* danke.

# Abbreviations

| | |
|---|---|
| **aa** | Amino acid |
| **APA** | Alternative polyadenylation |
| **ATM** | Ataxia-telangiectasia-mutated |
| **ATR** | Ataxia Telangiectasia and Rad3-related kinase |
| **bGH** | Bovine Growth Hormone |
| **BLAST** | Basic Local Alignment Search Tool |
| **bp** | Base pair(s) |
| **BSA** | Bovine Serum Albumin |
| **C-terminus** | Carboxyl-terminus |
| **CBC** | Cap-Binding Complex |
| **cDNA** | Complementary DNA |
| **CLAMP** | Chromatin-Linked Adaptor for Male-specific lethal [MSL] Proteins |
| **CPA** | Cleavage and polyadenylation |
| **CRAS** | Coding Region Activating Sequence |
| **CTD** | Carboxyl-Terminal Domain |
| **CTIF** | CBP80/20-dependent Translation Initiation Factor |
| **DAPI** | 4', 6-diamidino-2-phenylindole |
| **DMEM** | Dulbecco's Modified Eagle's Medium |
| **DMSO** | Dimethylsulfoxide |
| **DNA** | Deoxyribonucleic acid |
| **DNA-PK** | DNA-dependent protein kinase |
| **DSBs** | Double-stranded breaks |
| **dsDNA** | Double-stranded DNA |
| **DSE** | Downstream Sequence Element |
| **DTB** | Double Thymidine Block |
| **EDTA** | Ethylenediaminetetraacetic acid |
| **EdU** | 5-Ethynyl-2'-deoxyuridine |
| **EEF1A1** | Eukaryotic Translation Elongation Factor 1 Alpha 1 |
| **EGFP** | Enhanced green fluorescent protein |
| **EJC** | Exon-Junction Complex |
| **FACS** | Fluorescence-Activated Cell Sorting |
| **FBS** | Fetal Bovine Serum |
| ***g*** | G-force |

| | |
|---|---|
| **g** | Grams |
| **gDNA** | Genomic DNA |
| **gRNAs** | CRISPR-Cas9 guide RNAs |
| **GTFs** | General Transcription Factors |
| **HBSS** | Hanks' Balanced Salt Solution |
| **HDE** | Histone Downstream Element |
| **HeLa-FlpIn** | HeLa Flp-In T-Rex |
| **HLB** | Histone Locus Body |
| **hnRNP** | Heterogeneous nuclear RNP |
| **hr** | Hour(s) |
| **hTERT** | Human telomerase reverse transcriptase |
| **Kb** | Kilo bases |
| **kDa** | Kilodalton |
| **L** | Litre |
| **LB** | Luria-Bertrani broth |
| **M** | Molar |
| **MDa** | Mega Dalton |
| **mESCs** | Mouse embryonic stem cells |
| **min** | Minute(s) |
| **miPSCs** | Mouse induced pluripotent stem cells |
| **miRNA** | MicroRNA |
| **mRNA** | Messenger RNA |
| **MSCI** | Meiotic Sex Cromosome Inactivation |
| **NCBI** | National Center for Biotechnology Information |
| **ncRNA** | Non-coding RNA |
| **ng** | Nanogram |
| **NHEJ** | Non-Homologous End Joining |
| **nm** | Nanometer |
| **NMD** | Nonsense-Mediated Decay |
| **NPAT** | Nuclear Protein mapped to the AT locus |
| **NPCs** | Nuclear Pore Complexes |
| **NXF-1** | Nuclear Export Factor 1 |
| **ORF** | Open Reading Frame |
| **OTF-1** | Octamer-binding Franscription Factor 1 |
| **P-TEFb** | Positive Transcription Elongation Factor b |
| **PABP** | Poly(A)-Binding Protein |
| **PAS** | Polyadenylation site |
| **PBS** | Phosphate Buffered Saline |
| **PBS-CHX** | PBS containing 100 micro g/mL cycloheximide |
| **PCR** | Polymerase Chain Reaction |
| **PDDF** | Persistent DNA damage foci |
| **PFGE** | Pulsed-Field Gel Electrophoresis |

| | |
|---|---|
| **PI3K** | Phosphatidylinositol-3 Kinases |
| **PIC** | Pre-Initiation Complex |
| **Pol I** | RNA Polymerase I |
| **Pol II** | RNA Polymerase II |
| **Pol III** | RNA Polymerase III |
| **Pre-mRNA** | Precursor messenger RNA |
| **PTBP1** | Polypyrimidine tract-binding protein 1 |
| **RBPs** | RNA-Binding Proteins |
| **RD** | Replication-dependent |
| **RI** | Replication-independent |
| **RNA** | Ribonucleic acid |
| **RNAi** | RNA interference |
| **RNAseq** | RNA sequencing |
| **ROS** | Reactive oxigen species |
| **rpm** | Revolutions per minute |
| **rRNA** | Ribosomal RNA |
| **sec** | Second(s) |
| **SLIP1** | SLBP-Interacting Protein 1 |
| **snRNPs** | Small nuclear ribonucleoprotein particles |
| **SSCSs** | Subtype-Specific Consensus Sequences |
| **ssDNA** | Single-stranded DNA |
| **ssODN** | Signle-stranded donor oligonucleotide |
| **TES** | Transcription End Site |
| **Tet** | Tetracycline |
| **TREX** | Transcription Export Complex |
| **tRNA** | Transfer RNA |
| **TSS** | Transcription Start Site |
| **TUTs** | Terminal Uridylyl Transferases |
| **USE** | Upstream Sequence Element |
| **UTR** | Untranslated Region |
| **Y** | Pyrimidine |

# Contents

**Abstract** **I**

**Ackowledgements** **II**

**Abbreviations** **IV**

**Contents** **VII**

**List of Figures** **XI**

**List of Tables** **XIV**

**1 Introduction** **1**

1.1 Eukaryotic mRNA from birth to death: a quick overview . . . . . 1

1.1.1 Transcription and pre-mRNA maturation . . . . . 1

1.1.2 mRNA export . . . . . 4

1.1.3 mRNA translation . . . . . 5

1.1.4 mRNA decay . . . . . 6

1.2 Pre-mRNA 3'-end processing . . . . . 7

1.2.1 Cleavage and polyadenylation . . . . . 8

1.2.2 Histone mRNA 3'-end processing . . . . . 13

1.3 Histone genes . . . . . 17

1.3.1 Genomic organisation of histone genes . . . . . 17

1.3.2 Cell cycle regulation of histone genes . . . . . 18

1.3.3 Translation of replication-dependent histone mRNA . . . . . 20

1.3.4 Summary illustration: Life cycle of "bulk" polyadenylated mRNA *vs* RD histone mRNA . . . . . 22

1.3.5 Replication-independent histone genes . . . . . 23

1.4 Polyadenylation of replication-dependent histone mRNA . . . . . 24

1.5 H2AFX . . . . . 25

1.5.1 H2AFX gene . . . . . 25

1.5.2 H2A.X mRNA . . . . . 26

1.5.3 H2A.X protein . . . . . 27

1.6 Aim of this thesis . . . . . 31

**2 Materials and Methods** **33**
2.1 Cell culture . . . 33
2.1.1 Cell lines . . . 33
2.1.2 Culturing conditions . . . 33
2.1.3 Cell passaging . . . 34
2.1.4 Cell freezing and thawing . . . 34
2.1.5 Cell counting . . . 35
2.2 Cell cycle synchronisation . . . 35
2.2.1 Nocodazole block . . . 35
2.2.2 Aphidicolin block . . . 36
2.2.3 Double thymidine block . . . 36
2.2.4 RPE-1 cell cycle arrest by contact inhibition . . . 37
2.3 Flow cytometry . . . 38
2.3.1 Propidium iodide staining . . . 38
2.3.2 EdU labelling . . . 38
2.3.3 Flow cytometer . . . 39
2.4 Transient transfection . . . 40
2.4.1 siRNA transfection . . . 40
2.4.2 Plasmid transfection . . . 40
2.5 Cloning . . . 40
2.5.1 Software . . . 40
2.5.2 Transformation . . . 41
2.5.3 Colony PCR . . . 41
2.5.4 H2AFX 3'UTR reporter cloning for transient transfection . . . 42
2.5.5 H2AFX 3'UTR reporter cloning for stable integration into HeLa-FlpIn cells . . . 44
2.6 CRISPR-Cas9 genome engineering . . . 49
2.7 DNA purification . . . 52
2.7.1 DNA precipitation . . . 52
2.7.2 Genomic DNA extraction . . . 52
2.7.3 Plasmid DNA isolation . . . 53
2.8 Sequencing . . . 53
2.8.1 Sanger Sequencing . . . 53
2.8.2 Illumina sequencing . . . 53
2.9 Reverse transcription . . . 54
2.10 PCR . . . 55
2.11 Real-time PCR . . . 55
2.12 Agarose gel electrophoresis (DNA) . . . 55
2.13 RNA extraction . . . 56
2.14 Poly(A)$^+$ mRNA isolation . . . 57
2.15 RNA size selection by urea gel electrophoresis . . . 57
2.16 Northern blot . . . 58

2.16.1 Agarose northern blot . . . 58
2.16.2 Northern blot probe design . . . 61
2.16.3 Northern blot membrane stripping . . . 62
2.16.4 Northern blot densitometry . . . 62
2.17 Western blot . . . 62
2.18 Protein quantification . . . 64
2.18.1 Nuclear-cytoplasmic fractionation . . . 64
2.19 Immunofluorescence . . . 65
2.19.1 Image acquisition . . . 66
2.19.2 Image analysis . . . 66
2.20 Pulsed-field gel electrophoresis . . . 67
2.21 Polysome gradient fractionation . . . 67
2.21.1 Gradient preparation . . . 68
2.21.2 Sample preparation . . . 68

**3 Results Part I: Polyadenylation of RD histone mRNA 70**
3.1 RD histone genes express polyadenylated mRNA in HeLa cells . . . 70
3.2 H2A.X mRNA cell cycle regulation . . . 80

**4 Results Part II: H2A.X poly(A)$^+$ mRNA 86**
4.1 H2A.X poly(A)$^+$ mRNA depletion . . . 86
4.2 Effect of H2A.X poly(A)$^+$ mRNA depletion on DNA damage signalling . . . 91

**5 Results Part III: H2A.X SL mRNA 100**
5.1 H2A.X SL mRNA "knock out" . . . 100
5.2 Effect of H2A.X SL mRNA "knock out" on the cell cycle distribution of H2A.X poly(A)$^+$ mRNA . . . 105
5.3 Effect of H2A.X SL mRNA "knock out" on DNA damage signalling . . . 107

**6 Results Part IV: Regulatory elements in the H2A.X poly(A)$^+$ mRNA 114**

**7 Discussion 127**
7.1 HeLa cells express polyadenylated RD histone mRNA . . . 127
7.2 Cell cycle regulation of the H2A.X mRNA isoforms . . . 128
7.3 The H2A.X poly(A)$^+$ mRNA is required for efficient $\gamma$H2A.X signalling . . . 130
7.4 The H2A.X protein incorporated into the chromatin in S-phase is required for DNA damage signalling . . . 134
7.5 The SL may regulate H2A.X poly(A)$^+$ mRNA levels in S-phase and a conserved region close to the PAS is required for H2A.X poly(A)$^+$ mRNA expression . . . 137
7.6 General conclusions . . . 143
7.7 Future directions . . . 144

**References 146**

**Supplementary Figures** **162**

# List of Figures


1.1 Mammalian polyadenylation site . . . 8
1.2 CPA machinery . . . 12
1.3 Replication-dependent histone 3'-end processing sequence elements . . . 14
1.4 Replication-dependent histone 3'-end processing machinery . . . 17
1.5 Histone cluster 1 . . . 18
1.6 Summary illustration: Life cycle of "bulk" polyadenylated mRNA *vs* RD histone mRNA . . . 22
1.7 H2A.X mRNA isoforms . . . 26

2.1 Nocodazole synchronisation protocol . . . 35
2.2 Aphidicolin synchronisation protocol . . . 36
2.3 Double Thymidine synchronisation protocol . . . 37
2.4 Plasmid maps of the H2AFX 3'-UTR reporter . . . 43
2.5 Original plasmid employed for stable integration into the HeLa-FlpIn cell line 45
2.6 Reverse trancscription using SuperScript IV . . . 54
2.7 Agarose northern blot - transfer setup . . . 59

3.1 Cell cycle synchronisation for RNAseq analysis . . . 73
3.2 RNA fractionation steps for RNAseq experiment . . . 74
3.3 Example of RD histone genes (RNAseq) . . . 75
3.4 Example of RD histone genes expressing poly(A)$^+$ mRNA isoform (RNAseq) 76
3.5 RNAseq. H2AFX . . . 77
3.6 H2A.X mRNA cell cycle distribution in HeLa cells (DTB) . . . 83
3.7 H2A.X mRNA cell cycle distribution in HCT116 cells (DTB) . . . 84
3.8 H2A.X mRNA expression in RPE-1 cells . . . 85

4.1 H2A.X poly(A)$^+$ mRNA depletion in HeLa cells . . . 89
4.2 H2A.X poly(A)$^+$ mRNA depletion in RPE-1 cells . . . 90
4.3 H2A.X poly(A)$^+$ mRNA depletion and DNA damage signalling . . . 95
4.4 H2A.X poly(A)$^+$ mRNA depletion rescue experiment (northern) . . . 96
4.5 H2A.X poly(A)$^+$ mRNA depletion rescue experiment (IF) . . . 97
4.6 H2A.X mRNA decay and nucleocytoplasmic distribution after etoposide treatment. . . . 98
4.7 Effect of DNA damage on H2A.X mRNA translation . . . 99

5.1 HDE deletion constructs . . . 102
5.2 Sequenced gDNA of HeLa and HCT116 clones . . . 103
5.3 The H2AFX HDE deletion efficiently impedes H2A.X SL mRNA expression 104
5.4 Cell cycle distribution of the H2A.X poly(A)$^+$ mRNA in H2A.X SL mRNA "knock out" cells (dHDE) . . . 106
5.5 Effect of DNA damage on H2A.X mRNA isoforms in H2A.X SL mRNA "knock out" cells . . . 109
5.6 Effect of DNA damage on H2A.X protein levels in H2A.X SL mRNA "knock out" cells . . . 111
5.7 Effect of H2A.X SL mRNA "knock out" on $\gamma$H2A.X signalling in S-phase and non-S-phase cells(HeLa) . . . 112
5.8 Effect of H2A.X SL mRNA "knock out" on $\gamma$H2A.X signalling in S-phase and non-S-phase cells (HCT116) . . . 113

6.1 Transgene H2A.X pre-mRNA expressed in the different HeLa Flp-In T-REx cell lines . . . 116
6.2 PCR approach to distinguish transgene and endogenous H2AFX . . . 117
6.3 RT-PCR of asynchronously growing HeLa Flp-In T-REx cell lines . . . 121
6.4 RT-PCR of size selected RNA from asynchronously growing HeLa Flp-In T-REx cell lines . . . 122
6.5 PCR of DTB-synchronised HeLa Flp-In T-REx cell lines (Total RNA) . . . 123
6.6 PCR of DTB-synchronised HeLa Flp-In T-REx cell lines (poly(A)$^-$ and poly(A)$^+$ RNA - Total H2A.X primers) . . . 124
6.7 PCR of DTB-synchronised HeLa Flp-In T-REx cell lines (poly(A)$^-$ and poly(A)$^+$ RNA - poly(A)$^+$ H2A.X primers) . . . 125
6.8 Summary image from Figure 6.7 . . . 126

7.1 Proposed model for alternative H2A.X mRNA translation mechanisms . . . 133
7.2 Schematic summarising the effect of the depletion of either of the two H2A.X mRNA isoforms on $\gamma$H2A.X signalling . . . 136
7.3 Proposed model for H2A.X poly(A)$^+$ mRNA regulation by SLBP . . . 139
7.4 Proposed model for co-transcriptional regulation of the H2A.X poly(A)$^+$ mRNA by the conserved region . . . 141
7.5 Proposed model for post-transcriptional regulation of the H2A.X poly(A)$^+$ mRNA by the conserved region . . . 142
S1 Single cell FACS sorting in 96-well plate - gating strategy . . . 163
S2 Image projection macro for FIJI . . . 164
S3 Image quantification macro for FIJI . . . 165
S4 Flow cytometry of EdU-labelled HeLa cells synchronised by DTB . . . 166
S5 Flow cytometry of EdU-labelled RPE-1 cells depleted of H2A.X polyA RNA and synchronised by contact inhibition . . . 167
S6 Flow cytometry of EdU-labelled HeLa dHDE CRISPR cells synchronised by DTB . . . 168

S7 Flow cytometry of EdU-labelled HCT116 dHDE CRISPR cells synchronised by DTB . . . 169
S8 H2A.X poly(A)$^+$ mRNA depletion in HCT116 cells (total H2A.X protein levels) . . . 170
S9 H2A.X poly(A)$^+$ mRNA depletion in HCT116 cells ($\gamma$H2A.X signalling) . . 171
S10 gDNA PCR of the HeLa-FlpIn cell lines proving HeLa cells have two H2AFX genes . . . 172
S11 Minus-RT PCR of the HeLa-FlpIn cell lines synchronised by DTB . . . 172
S12 Flow cytometry of EdU-labelled HeLa-FlpIn cell lines synchronised by DTB 173
S13 PCR cycle number control . . . 174

# List of Tables


1.1 Human histone variants . . . . . 23

2.1 Cell lines . . . . . 33
2.2 siRNAs . . . . . 40
2.3 Primers employed for the H2AFX reporter cloning . . . . . 44
2.4 Primers employed for Sanger Sequencing of pcDNA5-FRT-H2AFX constructs 46
2.5 Primers employed for site-directed mutagenesis . . . . . 48
2.6 Primers to PCR-amplify H2A.X SL and poly(A) RNA . . . . . 49
2.7 CRISPR-Cas9 guide RNA and HR template sequences . . . . . 52
2.8 Agarose gel preparation protocol for northern blotting. . . . . 58
2.9 Northern blot probes . . . . . 61
2.10 Western blot primary antibodies. . . . . 64
2.11 Primary and secondary antibodies employed for immunofluorescence. . . . . 66
2.12 PFGE electrophoresis settings . . . . . 67

3.1 Histone genes in HeLa cells. . . . . 78

# Chapter 1

# Introduction

In the following chapter, the life cycle of "bulk" eukaryotic messenger RNA (mRNA) as well as the interconnection between the different steps is briefly described (Section 1.1), thereafter focusing on precursor-mRNA (pre-mRNA) 3'-end processing in Section 1.2 and then specifically on histone genes and the differences in the life cycle of replication-dependent histone mRNA (Section 1.3). Finally, specifically the H2AFX gene is described, as well as the functions of the H2A.X protein, in Section 1.5.

## 1.1 Eukaryotic mRNA from birth to death: a quick overview

### 1.1.1 Transcription and pre-mRNA maturation

Each gene can be considered a single transcription unit comprised of promoter, gene body and terminator [1]. In eukaryotes, there are three RNA polymerases [2]. RNA polymerase I (Pol I) transcribes ribosomal RNA genes [3,4], RNA polymerase III (Pol III) stable small non-coding RNA genes such as transfer RNA [5] and RNA polymerase II (Pol II) mainly protein-coding genes, giving rise to mRNA as well as unstable long non-coding RNA [6].

mRNA synthesis by Pol II is initiated at the gene promoter in collaboration with the general transcription factors (GTFs) TFIIB, TFIID, TFIIE, TFIIF and TFIIH, together forming the pre-initiation complex (PIC). GTFs, together with other associated factors, enable promoter recognition, separation of the two DNA strands and positioning of the Pol II active site onto the template DNA strand, allowing RNA synthesis to initiate from the

transcription start site (TSS) [7]. Most polymerases undergo promoter-proximal pausing about 30-60 nucleotides downstream of the TSS [8, 9]. This process is thought to be a critical regulatory step in the transcription cycle and can serve as quality checkpoint for pre-mRNA 5'-capping and Pol II modification before entering the productive elongation state [8, 10, 11]. However, several recent studies imply that paused Pol II may reflect a dead end process that may act to block productive Pol II initiation and may require termination of the paused transcript to allow productive transcription [12]. Pol II is released into productive elongation after its activation by positive transcription elongation factor b (P-TEFb). Additional elongation factors enhance Pol II processivity, chromatin remodelling and regulate co-transcriptional pre-mRNA processing [9].

As soon as the nascent pre-mRNA appears at the Pol II RNA exit channel, 5'-capping takes place [13]. This is the first co-transcriptional pre-mRNA processing reaction, where a 7-methyl guanosine is linked to the first nucleotide of the nascent RNA transcript via a 5'-5' triphosphate bridge [14]. The 5'-cap is bound by proteins of the cap-binding complex (CBC), specifically CBP20 and CBP80, in the nucleus and later by eIF4E in the cytoplasm. The resulting cap structure influences splicing [15, 16], pre-mRNA 3'-end processing [17], mRNA export [18, 19], translation [20] and protects the mRNA from 5' to 3' exonucleoysis [21].

The second step of co-transcriptional pre-mRNA maturation is splicing: the excision of introns from the nascent transcript [22]. Specific consensus sequences in the pre-mRNA mark the 5' and 3' splice sites that are recognised by specific small nuclear ribonucleoprotein particles (snRNPs) of the spliceosome. The major spliceosome consists of the U1, U2, U4/U6 and U5 snRNPs, whereas the minor spliceosome contains the U11, U12, U4atac/U6atac and U5 snRNPs. Mechanistically, both complexes act in a similar manner, with the main difference in the consensus sequences of the splice sites [23]. Introns spliced by one or the other are often called U2-type and U12-type introns, respectively. The detailed splicing mechanism is beyond the scope of this thesis and is described elsewhere [24]. Importantly, splicing has effects on gene regulation, including transcription, 3'-end processing, mRNA stability, export and translation [20, 25–28]. Furthermore, alternative splicing dramatically increases the protein repertoire [29, 30].

When Pol II reaches the end of the gene, the RNA transcript is cleaved and released

from the DNA template. The pre-mRNA 3'-end processing reaction of most eukaryotic mRNA consists of an endonucleolytic cleavage reaction followed by the addition of a non-templated poly(A)-tail to the resulting 3'-end. The exception is replication-dependent histone pre-mRNA, which is not polyadenylated, but instead ends in a 3' stem-loop structure. A detailed description of both these mechanisms follows in Section 1.2. The mature mRNA 3'-end not only protects the mRNA from degradation but also stimulates transcription termination, mRNA export and translation [20, 31, 32].

Even though the RNA transcript is released from the DNA template, Pol II remains bound to the DNA requiring further steps for transcription termination. When 3'-end processing signals appear in the nascent RNA, processing factors associate with the polymerase leading to Pol II pausing [33]. Furthermore, the nascent transcript frequently invades the DNA and anneals to the template strand forming an R-loop structure, that reinforces this pausing effect [34]. Once the mRNA is released from the DNA template, Pol II continues transcribing until the 5' to 3' exonuclease Xrn2 degrades the 3' cleavage product. When Xrn2 catches up with Pol II it induces conformational changes in the active center of the polymerase leading to its release from the DNA template hence terminating transcription [35, 36]. Xrn2 does not seem to terminate Pol II at histone genes [35, 37], where CPSF-73 seems to act instead [38, 39].

### Co-transcriptional mRNA processing and mRNP formation: importance of the Pol II CTD

The first evidence of co-transcriptional pre-mRNA 3'-end processing was provided in 1986, where a mutated polyadenylation signal abolished 3'-end pre-mRNA processing as well as transcription termination [31]. Around the same time, it was shown that the exchange of a Pol II promoter for a Pol I or Pol III promoter negatively influenced downstream processing events of the resulting transcript, including polyadenylation, stability, nucleocytoplasmic export as well as splicing [40, 41]. The Pol II carboxyl-terminal domain (CTD) has been shown to be crucial for the different pre-mRNA maturation steps (see below). The absence of such domain in Pol I and Pol III may explain the difference in the resulting transcripts (*e.g.* they are not polyadenylated). Since these early discoveries, numerous studies have provided evidence for the tight interconnection and coordination of the different steps

in the mRNA life cycle and it is now clear that these different processes influence one another's specificity and efficiency (for review see [13, 42–47]).

During the different mRNA maturation steps, the transcript is coated with proteins, giving rise to messenger ribonucleoprotein complexes (mRNPs) ready to be exported to the cytoplasm. A failure in any of these processing steps leads to the formation of an incompetent mRNP that is then targeted for decay in the nucleus to avoid export and translation of misprocessed mRNA, which could spell disaster for the cell [48–50].

Both co-transcriptional pre-mRNA processing and mRNP formation are mainly directed by Pol II, specifically by the carboxyl-terminal domain (CTD) of the largest of its 12 subunits [51]. The CTD is a heptapeptide repeat of the consensus sequence, $Y_1S_2P_3T_4S_5P_6S_7$, ranging from 26 in budding yeast to 52 in mammals, that is subjected to extensive reversible posttranslational modifications, especially phosphorylation [52]. Particular CTD phosphorylation patterns recruit specific protein factors, involved in regulating the transcription cycle and the different steps of pre-mRNA maturation. Recently, a "CTD code" for human genes has been established: unphosphorylated CTD as well as $S_5$ phosphorylation predominate at the TSS. $S_5$ phosphorylated CTD also marks actively spliced exons, whereas $S_2$ phosphorylation increases towards the transcription end site (TES) [33]. In line with this, truncated CTD leads to a failure in 5' capping, splicing and 3'-end processing [53]. The following examples explain these results: 5' capping enzymes as well as the spliceosome directly bind to $S_5$ phosphorylated CTD [22, 54]. Similarly, cleavage and polyadenylation factors, *e.g.* CstF-50 and Pcf-11, also associate with the Pol II CTD [55, 56]. Other examples are reviewed in [44, 46, 52, 57, 58].

### 1.1.2 mRNA export

The now mature mRNP has to be exported from the nuclear compartment to the cytoplasm where protein translation and eventually mRNA degradation takes place. The nuclear envelope comprises a double lipid bilayer that is interrupted by transport channels called nuclear pore complexes (NPCs) that allow continous exchange of molecules and macromolecules between the nucleus and the cytoplasm. Each human NPC has a molecular mass of ∼120 MDa and is composed of multiple copies of ∼30 different nucleoporins. Those facing the nuclear site form the so-called "nuclear basket" while those facing the cytoplasm

are assembled into "cytoplasmic filaments" (structural details are reviewed in [59]). The nucleoporins at the interior of the NPC have disorderd protein sequences enriched in phenylalanine-glycine (FG) repeats forming a hydrophobic hydrogel that restricts passive diffusion of molecules larger than 30 kDa [60, 61]. Larger molecules require chaperones for efficient transport through the permeability barrier of the NPC.

The transport factor facilitating the export of mRNPs is the heterodimeric nuclear export factor NXF-1·NXT-1 (also called TAP-p15). NXF-1·NXT-1 acts as molecular signature that marks mRNA as "export-competent" when deposited onto the mRNA with help of multiple adaptor proteins, such as transcription export complex (TREX) [62–65]. TREX is recruited to a region close to the mRNA 5'-end, in a cap- and splicing-dependent manner [66]. Furthermore, splicing regulator SR proteins and components of the EJC can also recruit NXF-1 to correctly spliced mRNA [28, 67]. NXF-1·NXT-1 directly interacts with nucleoporin FG-repeats allowing mRNP transport throught the NPC. When exiting on the cytoplasmic face, cytoplasmic filament nucleoporins activate the DEAD-box helicase DDX19 which removes NXF-1·NXT-1 from the mRNP, releasing it into the cytoplasm [68]. These series of events ensure that mRNP transport is unidirectional only from the nucleus to the cytoplasm.

### 1.1.3 mRNA translation

The mRNPs released from the NPC into the cytoplasm are bound by the CBC at their 5'-end, the shuttling nuclear poly(A)-binding protein (PABPN1) coating their 3' poly(A)-tail and by the exon-junction complex (EJC) deposited during splicing [20]. These mRNPs diffuse throughout the cytoplasm until they encounter the translation machinery [69]. The first round of translation is referred to as "pioneer round of translation" and is directed by the CBC proteins CBP20 and CBP80 [20, 70]. During the pioneer round of translation, one or more ribosomes displace the RNA-binding proteins (RBPs) that coat the ORF and 5' UTR in the mRNA, including the EJCs [69, 71, 72]. This process adds an important quality control level. *E.g.*, EJCs stimulate the degradation of transcripts that contain a premature stop codon in a process called nonsense-mediated decay (NMD) [73]. After the pioneer rounds of translation, CBC is replaced by eIF4E and PABPN1 by its cytoplasmic counterpart PABPC1 [74] that support steady-state mRNA translation.

Efficient translation initiation may be facilitated by circularisation of the mRNA through molecular contacts between the mRNA 3'- and 5'-ends. The resulting "closed-loop" conformation is established by protein-protein interactions between eIF4E and PABPC1, through eIF4G [75–81].

Mechanistically, the pioneer round of translation and steady-state translation are similar [20]. mRNA translation can be divided into initiation, elongation, termination and ribosome recycling. During translation initiation, the 43S preinitiation complex containing the 40S ribosomal subunit, the methionyl-tRNA and translation initiation factors (eIFs), binds the mRNA 5'-end and scans along the transcript until it encounters the AUG codon complementary to the methionyl-tRNA anticodon [82]. Then the 60S ribosomal subunit binds forming the 80S ribosome which starts moving processively along the mRNA, adding a new amino acid per codon to the nascent protein through the action of aminoacyl-tRNAs and eukaryotic translation elongation factors (eEFs). Once the ribosome reaches the stop codon, the nascent protein is released by eukaryotic peptide chain release factors (eRFs) and the ribosomes are recycled as separate 40S and 60S subunits. The details and regulation of this mechanism are reviewed elswhere [77, 83–87].

### 1.1.4 mRNA decay

Regulating RNA stability is important at multiple levels. In the nucleus, misprocessed mRNA must be degraded to avoid its export to the cytoplasm and subsequent translation into deleterious proteins [88]. Similarly, most ncRNA is degraded before release into the nucleoplasm [89]. In the cytoplasm, mRNA levels must be regulated to control protein expression levels, and mRNA decay is activated as part of quality control mechanisms (*e.g.* NMD).

mRNA can be degraded by exonucleolysis from the 5'- or the 3'-end. In the nucleus, 5'-3' decay only represents a minor pathway, starting with the removal of the cap structure by decapping enzymes, such as DCP2, followed by 5'-3' exonucleolysis by XRN2 [90–92]. The major RNA turnover pathway in the nucleus is 3'-5' degradation by the nuclear exosome, which in humans contains the catalytic subunits DIS3 and RRP6 (EXOSC10). RRP6 is enriched in the nucleolus possibly being the only catalytic subunit of the exosome in this compartment, whereas DIS3 predominates in the nucleoplasm [93].

In the cytoplasm, the major mRNA decay pathway is the so-called "deadenylation-dependent decapping pathway". It is initiated by removal of the poly(A)-tail by 3'-5' poly(A)-specific exoribonucleases, specifically the CCR4-NOT and PAN2-PAN3 complexes as well as PARN in mammals [94–98]. Deadenylation is followed by decapping, likely related to mRNP circularity. Degradation then initiates from the 5'-end by the DCP1-DCP2 complex and XRN1 exonuclease, respectively [88, 99, 100]. Alternatively or additionally, shortening of the poly(A)-tail is followed by 3'-5' degradation by the cytoplasmic exosome, which contains the strictly cytoplasmic catalytic subunit DIS3L [97, 101, 102]. Relatively recently, it has been shown that cytoplasmic deadenylation can be bypassed by uridylation by terminal uridylyl transferases (TUTs). These enzymes catalyse the addition of UTP to the mRNA 3'-end that subsequently recruits the decapping machinery [98, 102–105]. A recent publication shows that most of the U-tails are added to short poly(A)-tails (shorter than 25 nucleotides) [106].

## 1.2 Pre-mRNA 3'-end processing

In eukaryotes, co-transcriptional 3'-end processing is an essential step in pre-mRNA maturation, influencing multiple steps in the mRNA life cycle, including transcription termination, mRNA stability, export and translation [32, 42–44, 46, 107, 108], as briefly mentioned above.

Most eukaryotic mRNA are processed by cleavage and polyadenylation (CPA), replication-dependent histone mRNA being the only coding mRNA processed by a different mechanism dependent on the U7 small nuclear ribonuclear protein particle (U7-dependent processing). Both mechanisms depend on *cis*-elements in the pre-mRNA that are recognised by specific *trans*-acting factors, defining the processing site and catalysing the 3'-end processing reaction, respectively. Most eukaryotic mRNA contains more than one 3'-end processing site, generating different mRNA or even different protein isoforms, by a process called alternative polyadenylation (APA). However, as will be presented in this thesis, some histone mRNA can also be processed at alternative 3'-end processing sites. Correct 3'-end processing site choice is essential, as differences in sequence or length of the 3'-untranslated region (3'-UTR) can influence mRNA fate due to the presence of different binding sites for RNA-binding proteins or regulatory ncRNA, such as RNA (miRNA) [109–115].

Distinguishing between these two 5'-end processing mechanisms is of crucial importance to the results presented in this thesis and are hence described in detail in the following two Subsections (1.2.1 and 1.2.2).

### 1.2.1 Cleavage and polyadenylation

#### *Cis*-acting elements

The CPA reaction efficiency is determined by well-defined sequence elements in the pre-mRNA. In mammalian cells there are four main sequence elements that define the core polyadenylation site (PAS) in the 3'-UTR of pre-mRNA: a highly conserved hexanucleotide AAUAAA, a less conserved U- or GU-rich downstream sequence element (DSE), UGUA or a U-rich upstream sequence element (USE) and the CA dinucleotide marking the actual cleavage site (Figure 1.1). Additional sequence elements can be recognized in some cases, frequently related to APA.

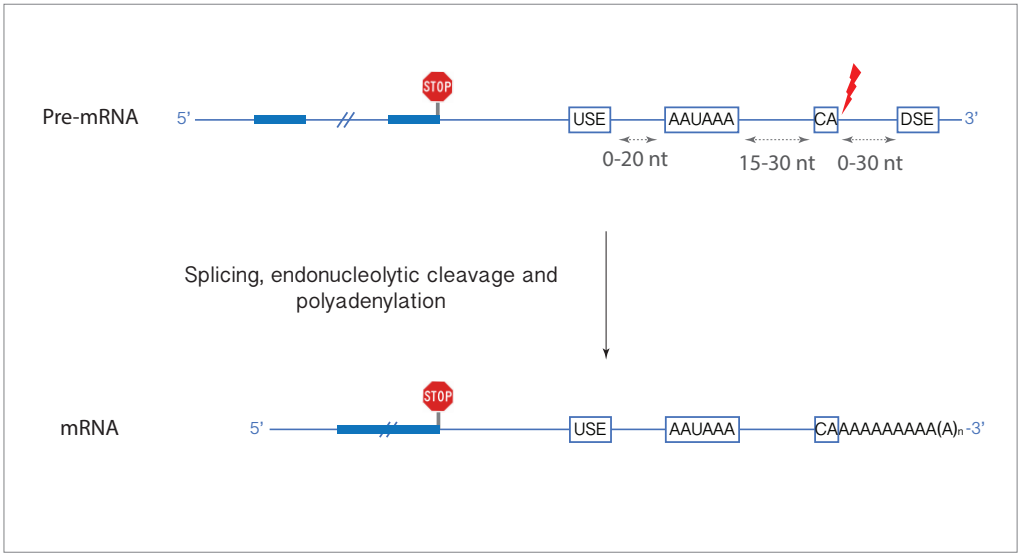



**Figure 1.1 Sequence elements defining the mammalian polyadenylation site.**

The AAUAAA consensus sequence was discovered by Proudfoot and Brownlee in 1976 [116]. It is situated 15-30 nucleotides upstream of the cleavage site and essential for CPA. Since its discovery, several variations of this sequence have been discovered, the most frequent being AUUAAA [117]. Changes in the consensus sequence lead to different binding affinity of the corresponding *trans*-acting factor and are thus frequently associated

with tissue-specific regulation or APA [32, 117, 118]. Mutations of the consensus sequence in two or more nucleotides can abolish PAS recognition and hence lead to inefficient gene expression and ultimately to disease [118–120].

Cleavage site definition is also influenced by the USE and the DSE. The USE consensus sequence is UGUA, frequently being present in more than one copy [121, 122]. Additional USEs include U-rich sequences. The DSE is located in a region 30 nucleotides downstream of the cleavage site [123] and is less conserved. It can include a U-rich element, the GU-rich consensus YGUGUUYY, where Y is a pyrimidine, or both [118, 124, 125]. The DSE enhances 3'-end processing efficiency and can influence the cleavage position but is inessential for CPA [124, 126–128].

Cleavage occurs 3' of a CA dinucleotide. The cleavage downstream of this CA dinucleotide is conserved among eukaryotes and mutations in these nucleotides can have severe consequences on gene expression, leading to disease [118].

### *Trans*-acting factors

The *cis*-elements described above execute their function *via* specific *trans*-acting factors. These include multiple highly conserved protein factors, frequently assembled in subcomplexes [129]: cleavage and polyadenylation specificity factor (CPSF), cleavage stimulation factor (CstF), mammalian cleavage factor I (CFIm), mammalian cleavage factor II (CFIIm), poly(A) polymerase (PAP), symplekin and Pol II [118, 123, 130] (Figure 1.2).

CPSF is a large protein complex composed of six subunits: CPSF-160, CPSF-100, CPSF-73, CPSF-30, WDR-33 and Fip1 (except for WDR-33 and Fip1, the numbers indicate the molecular weight of the mammalian factor in kDa). For a long time, the CPSF-160 subunit of CPSF was thought to recognise the AAUAAA hexanucleotide. However, recent studies have shown that CPSF-30 and WDR-33 are the direct binding contacts between the CPSF complex and AAUAAA, CPSF-160 rather acting as an essential scaffold for the binding of CPSF-30 and WDR-33 to the hexanucleotide [131–134]. CPSF-73 is a zinc-dependent hydrolase of the metallo-$\beta$-lactamase superfamily and was shown to be the endonuclease catalysing pre-mRNA cleavage. However, CPSF-73 shows weak enzymatic activity in isolation and requires the other CPSF components for full activity [135, 136].

CPSF-100 is structurally similar to CPSF-73 but it is catalytically inactive and its function is still to be determined [136]. Furthermore, Fip1 binds U-rich sequences upstream of AAUAAA as well as PAP, stimulating its activity and enabling poly(A) addition after the cleavage [137]. Additionally, Fip1 was shown to influence stem-cell maintenance by regulating APA [138].

CstF is a hexamer composed of two copies of each subunit: CstF-77, CstF-64 and CstF-50 [139]. The RRM domain of CstF-64 is responsible for CstF binding to the GU-rich DSE, contributing to PAS definition [139,140]. In agreement, CstF-64 has been implicated in regulating tissue-specific APA [141,142]. Alternative CstF-64 protein isoforms contribute to APA allowing the formation of CstF complexes with different sequence recognition properties [143]. Specifically, there are several alternatively spliced Cstf-64 isoforms and, at least in human and mouse, a paralog named CstF-64Tau. The latter was originally discovered in male germ cells, where it takes over CstF-64 function (the canonical CstF-64 gene is located on the X-chromosome and thus not expressed in testes, whereas CstF-64Tau is located on chromosome 19) [144,145]. More recently, CstF-64Tau has been shown to be ubiquitously expressed and able to compensate for CstF-64 loss in CPA [146]. The CstF-50 subunit directly tethers the CstF complex to Pol II CTD, stimulating co-transcriptional 3'-end processing [55]. CstF-77 is the bridging factor between CstF-64 and CstF-50 and it interacts with CPSF-160, stabilizing the complex on the pre-mRNA [139].

CFIm is composed of four subunits: CFIm-72, CFIm-68, CFIm-59 and CFIm-25. In its active form, CFIm is a heterotetramer composed of two CFIm-25 subunits and two of the larger subunits. The CFIm-25/CFIm-68 heterotetramer specifically recognises the USE consensus sequence UGUA through direct binding of the CFIm-25. CFIm-68 enhances CFIm-25 binding to the pre-mRNA [121]. Additionally, it has been shown that CFIm can bind two UGUA elements simultaneously, forming a RNA loop structure [122]. CFIm has been implicated in regulating APA [147,148] and it was shown to be recruited by the splicing machinery to stimulate 3'-end processing [26].

CFIIm has two subunits: hPcf11 and hClp1, which interact with each other and with CFIm and CPSF [149,150]. CFIIm may contribute to polyadenylation site recognition by binding to far downstream G-rich sequence elements [**?**]Schafer2018). Moreover, Pcf11 interacts with Pol II CTD and facilitates mRNA export [151] as well as 3'-end processing

and transcription termination [152].

Another important protein involved in the 3'-end processing reaction is the scaffold-like protein symplekin. This is suggested to act as a platform for the assembly of the 3'-end processing machinery [153] interacting with CPSF and CstF complexes through CPSF-100, CPSF-73 and CstF-64 [146].

The enzyme responsible for the poly(A) addition is PAP [154]. It was shown to be recruited to the 3'-end processing machinery, by interacting with CPSF, where it adds a tail of ∼250 adenosine nucleotides to the resulting 3'-end of the newly synthesised mRNA [155,156]. The cytoplasmic poly(A) binding protein (PABPC) coating the poly(A)-tail protects the mRNA from 3'-5' degradation and facilitates translation initiation through its interaction with eIF4G [157, 158]. The poly(A)-tail is gradually shortened during the mRNA life cycle by deadenylases, but it can also be re-extended by cytoplasmic poly(A) polymerases, modulating mRNA stability and translatability [97, 159, 160]. A recent study suggested that the average steady-state poly(A)-tail length is about 50-100 nucleotides in mammalian cells, corresponding to the intermediate length resulting from deadenylation [158].

Taken together, the CPA 3'-end processing machinery is highly complex and the molecular details of how the complexes assemble together on the pre-mRNA to eventually promote cleavage and polyadenylation are only recently beginning to be understood [133, 134]. As mentioned above, Pol II plays a crucial role in this process as many 3'-end processing factors are recruited to the CTD of elongating Pol II, positioning the factors into close proximity with the PAS as it emerges from the Pol II exit channel, increasing binding probability and efficiency. See Figure 1.2 for a schematic illustration of the mammalian CPA.

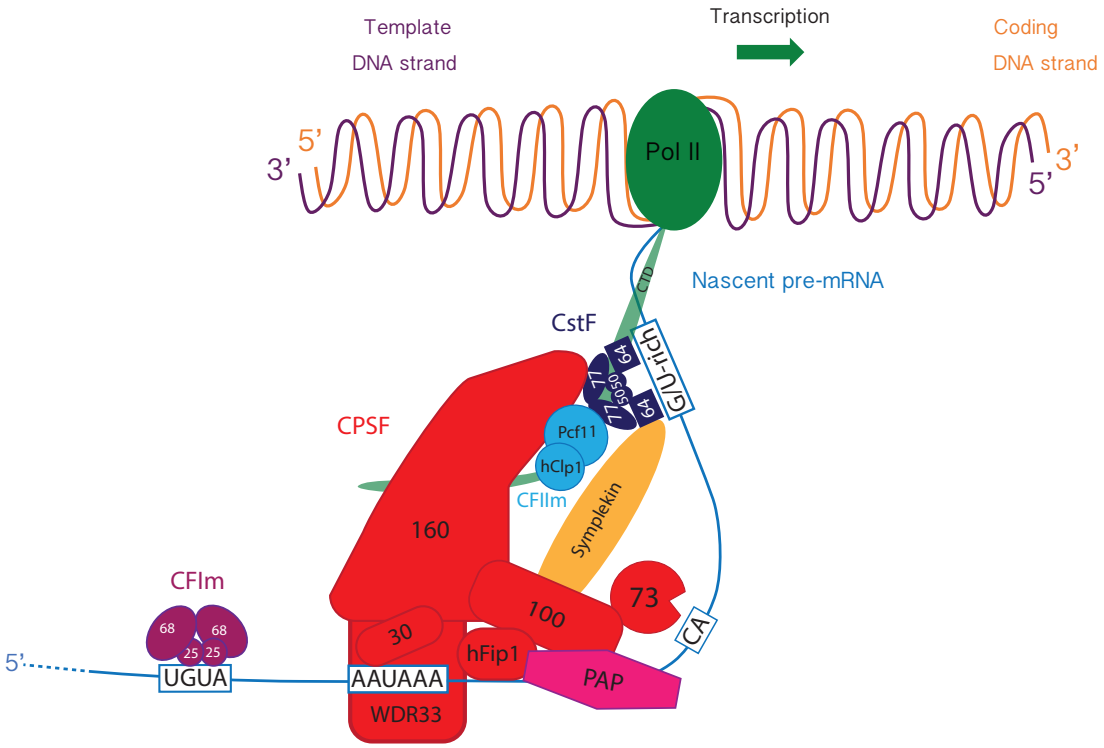



**Figure 1.2 Mammalian CPA machinery.** Mammalian cleavage factor I (CFIm) is a heterotetramer composed of two CFIm-25 subunits and two of the larger subunits (in this illustration CFIm-68), which recognises the consensus sequence UGUA. Cleavage and polyadenylation specificity factor (CPSF) is composed of CPSF-160, CPSF-100, CPSF-73, CPSF-30, WDR-33 and Fip1 and is responsible for the recognition of the AAUAAA hexamer (specifically by CPSF-30 and WDR33) and for exonucleolytic cleavage 3' of the CA dinucleotide by CPSF-73. Mammalian cleavage factor II (CFIIm) has two subunits: hPcf11 and hClp, and may contribute to poly(A) site definition by binding to far downstream G-rich sequence elements. Symplekin is a scaffold-like protein required for the assembly of the CPA machinery through protein-protein interactions. Cleavage stimulation factor (CstF) is a heterohexamer containing two of each subunits: CstF-77, CstF-64 and CstF-50, which recognises G/U-rich downstream sequence elements. Poly(A) polymerase (PAP) is recruited to the 3'-end processing machinery through CPSF and is the enzyme responsible for adding the poly(A)-tail to the resulting 3'-end.

### 1.2.2 Histone mRNA 3'-end processing

Histone proteins are essential for DNA packaging into nucleosomes to form chromatin. When the cell enters the S-phase of the cell cycle and starts DNA replication, the required amount of histone proteins has to be available for the assembly of nascent chromatin. Yet excess histone proteins are deleterious for the cell, leading to genomic instability and cytotoxicity, mainly caused by unspecific binding of positively charged histones to negatively charged molecules, such as DNA and RNA [161, 162]. Therefore, histone protein production is tighly coupled to DNA replication, increasing as the cell enters S-phase, kept high during S-phase and drastically decreased once DNA replication is concluded.

In metazoans there are five canonical histone proteins: four core histones, H2A, H2B, H3 and H4, and a linker histone, H1. Each nucleosome contains an octamer composed of two of each core histones, around which ∼147 bp DNA is wrapped [163]. The H1 linker histone stabilises the nucleosome and allows higher-order structuring to form chromatin [164]. Canonical histones are encoded by a family of replication-dependent (RD) histone genes, each canonical histone protein being encoded by multiple genes. These genes are intronless and physically linked in large clusters. Transcription is performed by Pol II and the resulting mRNA is 5' capped but not polyadenylated, requiring only a single 3'-end cleavage event to form mature mRNA [165, 166].

Coupling of RD histone gene expression to DNA replication is regulated at different levels: transcription increases 3- to 5-fold as the cell enters S-phase and, more importantly, the 3'-end processing reaction is 8- to 10-fold more efficient in S-phase, ensuring fast accumulation of mature RD histone mRNA. This effect is achieved by the unique 3'-end structure of RD histone mRNA, *i.e.* the unique *cis*-elements and their corresponding *trans*-acting factors, which play a role in the entire life cycle of RD histone mRNA, including export, translation and specific mRNA decay at the end of S-phase.

#### *Cis*-acting elements

Metazoan RD histone mRNA contains two main *cis*-elements that are essential for 3'-end formation: a highly conserved stem-loop structure (SL) positioned 5' to the cleavage site, and a purine-rich sequence called histone downstream element (HDE) situated downstream

of the cleavage site (see Figure 1.3).

The SL consists of a 16-nucleotide sequence (6-nucleotide stem and 4-nucleotide loop), localised about 20-75 nucleotides downstream of the stop codon. The SL consensus sequence is 5'-GGCUCUUUUCAGAGCC-3' for H1, H2B, H2A and H3 mRNA and 5'-GGCCCUUUUCAGGGCC-3' for H4 mRNA, with variations mainly at position 10 (C to A or T) ( [166] and own observation). CCAAA is the consensus before the stem-loop, and is essential for SLBP binding [167]. Mutations in several nucleotides of the SL or in the upstream consensus sequence lead to a severe decrease in processing efficiency [168]. The stem-loop is followed by a conserved AC-rich sequence with the consensus ACCCA in vertebrates [169]. Endonucleolytic cleavage takes place immediately downstream of this AC-rich element [170].

The HDE is a less conserved purine-rich element that in vertebrates has the consensus 5'-AAAGAGCTGT-3'. It is localised 9-12 nucleotides downstream of the AC-rich element. In mammals, this distance is essential for the definition of the cleavage position. Moving the HDE further downstream leads to a shift of the cleavage position further downstream and to a decrease in 3'-end processing efficiency [170]. This is possibly true for other vertebrates but not invertebrates, *e.g.* in *Drosophila* cleavage is always 4 nucleotides after the stemloop.

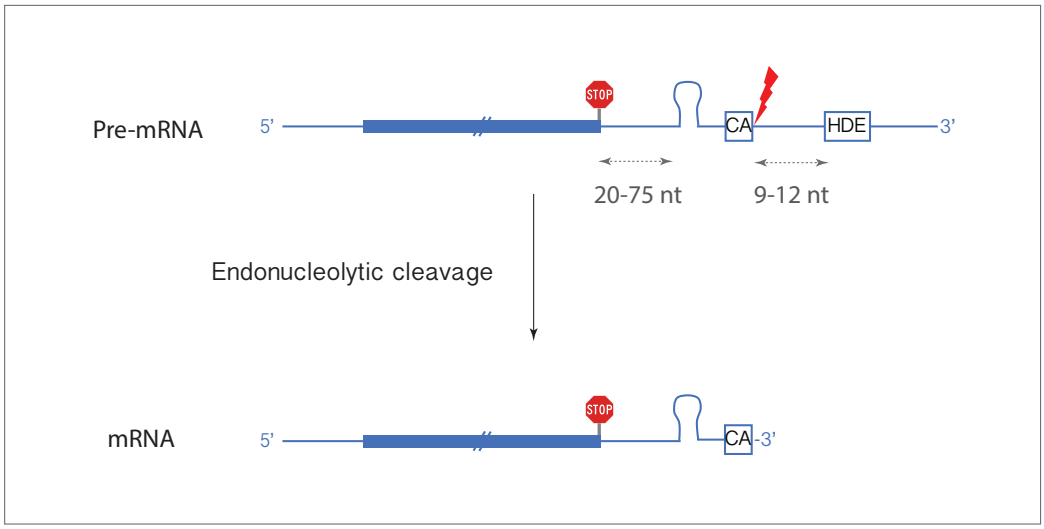



**Figure 1.3 Sequence elements defining the mammalian U7-dependent processing site typical of replication-dependent histone mRNAs.**

### *Trans*-acting factors

The main *trans*-acting factors in RD histone pre-mRNA 3'-end formation are: stem-loop binding protein (SLBP), U7 small-nuclear ribonucleoprotein particle (U7 snRNP), FLICE-associated huge protein (FLASH) and the histone cleavage complex (HCC) comprised of CPSF-73, CstF-64, CPSF-100 and symplekin, which are also present in the CPA machinery.

SLBP specifically recognises the conserved SL structure [171–174], binding 5' of the SL and requiring the 5 nucleotides immediately upstream of the SL [167, 175]. Recently, the crystal structure of SLBP bound to the SL was elucidated [167]. Interestingly, this study showed that SLBP mainly recognises the shape of the SL structure, only directly recognising one nucleotide: the second nucleotide of the stem, which corresponds to a G (see consensus sequences above) [167]. SLBP is essential for RD histone pre-mRNA 3'-end processing, as its depletion or the mutation of its RNA binding domain abolish the reaction [167, 168, 176]. One explanation for why SLBP is an essential factor is that it is important for U7 snRNP recruitment to the HDE [168, 177], which is also required for 3'-end formation [178], as described below. SLBP remains bound to the mRNA during its entire life cycle and is required in multiple aspects of RD histone mRNA metabolism, including mRNA export [179], translation [180–182] and stability [167].

The U7 snRNP contains a ∼70 nt RNA moiety with a 2,2,7-trimethylguanosine 5' cap structure and a ring of seven Sm-proteins. Two of these Sm-proteins are different from the Sm-proteins found in other U snRNPs: the typical Sm proteins D1 and D2 are substituted by the U7 snRNP-specific Lsm10 and Lsm11 [183]. The U7 snRNP binds to the HDE through direct base-pairing of the U7 snRNA 5'-end, thereby bringing the protein ring into close proximity of the mRNA cleavage site (see Figure 1.4). This binding is stabilised by SLBP and FLASH and is essential for histone mRNA processing [38, 177, 178]. It has been demonstrated that it is the Lsm11 protein of the U7 snRNP that interacts with the scaffold protein FLASH through its extended N-terminus [184, 185], serving as platform for the recruitment of the HCC that contains the endonuclease CPSF-73 [177]. Mutation of either Lsm10 or Lsm11 in *Drosophila* lead to impaired RD histone pre-mRNA 3'-end processing and severe developmental defects, proving both are essential for U7 snRNP function in RD histone mRNA 3'-end processing [186].

The HCC contains four protein factors: CPSF-73, CstF-64, CPSF-100 and symplekin [187]. The endonucleolytic activity of the HCC is provided by CPSF-73. It has been shown that CPSF-73 not only catalyses the endonucleolytic cleavage reaction but also the 5'-3' exonucleolysis of the resulting 3' cleavage product which contains a 5' phosphate, potentially promoting transcription termination [38, 39]. Immediate transcription termination promoted by CPSF-73 would explain the fast drop of Pol II occupancy downstream of the U7-dependent processing site observed for RD histone genes [188]. This possiblity is also supported by the fact that Xrn2 inhibition does not affect Pol II termination at histone genes [35, 37].

HCC assembly is mediated by the scaffold-like protein Symplekin, as its depletion impairs HCC formation [146]. Early studies detected a heat-labile factor which was invovled in cell-cycle regulation of processing [189], which was subsequently shown to be symplekin [190]. CstF-64 directly interacts with symplekin and FLASH and/or Lsm11, not only playing a role in HCC assembly but also in HCC recruitment to the platform created by FLASH and Lsm11 [146, 191]. Additionally, CstF-64 was suggested to be a cell-cycle regulated component of the HCC, directing the assembly of the factors usually associated with the CPA machinery to form the HCC instead [191]. This study also shows that depletion of CstF-64 leads to deficient RD histone mRNA 3'-end formation [191]. The role of CPSF-100 in this process is less clear but it is an essential part of the HCC [187, 192].

The fact that isolated U7 snRNP contains the HCC components suggests an association of these factors prior to pre-mRNA binding, to form so-called holo-U7 snRNP. FLASH associates to the U7 snRNP through an interaction with Lsm11 forming a platform for the recruitment of HCC. This entire complex then binds the RD pre-mRNA to catalyse the endonucleolytic cleavage giving rise to the mature mRNA [187, 193].

Recently, the novel endonuclease MBLAC1 has been implicated in an alternative U7 snRNP-independent mechanism to form the RD histone mRNA 3'-end [194]. However, this enzyme is not associated with the HCC [187].

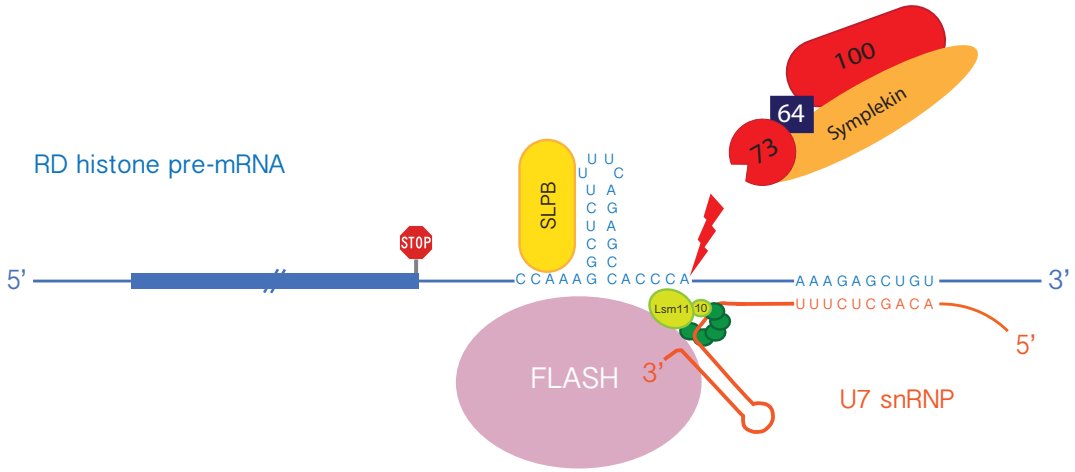



**Figure 1.4 Replication-dependent histone 3'-end processing machinery.** Eventhough the HCC containing CPSF-73, CstF-64, Symplekin and CPSF-100 are represented separately of the U7 snRNP-FLASH platform, it is possible that they form a complex together before binding to the pre-mRNA.

## 1.3 Histone genes

### 1.3.1 Genomic organisation of histone genes

RD histone genes are physically linked in the genome. In humans there is a major cluster on chromosome 6 (6p21–p22) containing 55 genes (Figure 1.5), and two minor clusters on chromosome 1 (1q21 and 1q42) containing 10 and 3 genes. RD histone gene nomenclature starts with the name of the corresponding cluster: HIST1, HIST2 and HIST3, respectively, followed by the histone type and a letter (*e.g.* HIST1H1C is one of the H1 linker histones in histone cluster 1). Additionally, there is one H4 gene located on chromosome 12 called HIST4H4 [195].

The physical proximity of the multiple RD histone genes may promote coordinated gene expression, ensuring stoichiometric production of large amounts of canonical histone proteins in the S-phase of the cell cycle. The molecular basis for this coordinated regulation lies in the assembly of the so-called histone locus body (HLB) at the histone loci [196, 197].

The HLB is a concentration of factors that are essential for RD histone gene expression at the histone cluster. It is present throughout interphase of the cell cycle [198–200] and contains the U7 snRNP as well as FLASH [184]. Interestingly, studies in *Drosophila*

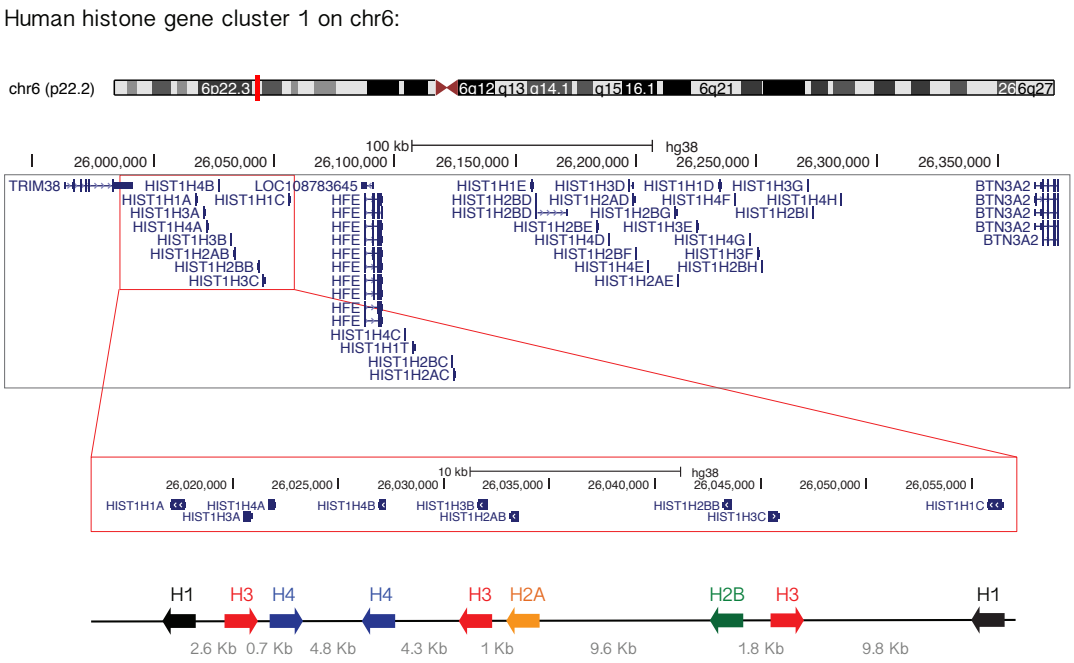



**Figure 1.5** Schematic illustration of histone cluster 1. The red box shows an amplification of part of the cluster. The histone genes are represented schematically bellow with the distances between the genes indicated in gray. The chromosome ideogram and the genes were taken from the NCBI Genome Browser.

showed that U7 snRNP and FLASH recruitment to form the HLB does not depend on the *cis*-elements on the RD histone pre-mRNA but rather on a genomic region within the histone cluster [201]. Specifically, a GA repeat *cis*-element within the bidirectional H3/H4 promoter [202]. Another factor involved in HLB formation and RD histone protein expression is the scaffold-like transcription factor Nuclear Protein mapped to the AT locus (NPAT), which is also present at the HLB throughout the cell cycle [197, 203]. There are other factors specifically recruited to the HLB in S-phase, such as the HCC [200], but also general factors involved in transcription, *e.g.* Pol II [197]. The HCC is only recruited to the U7 snRNP in S-phase in the HLB (oral communication by William Marzluff).

### 1.3.2 Cell cycle regulation of histone genes

The cell cycle regulation of RD histone gene expression occurs at multiple levels: transcription, pre-mRNA 3'-end processing efficiency, mRNA stability, as well as histone protein stability. When the cell approaches the G1/S-phase border, RD histone gene transcription increases (3- to 5-fold) and all synthesised pre-mRNA is rapidly processed

through an increase in 3'-end processing efficiency (8- to 10-fold). Consequently, there is a fast accumulation of mature RD histone mRNA (about 35-fold), which are efficiently exported and translated into canonical histone proteins that are then imported into the nucleus and incorporated behind the replication fork by specific chaperones [204–208]. Once DNA replication is concluded RD histone mRNA and excess histone proteins are rapidly degraded [209, 210]. The mechanism for excess histone degradation has been described in yeast, where it involves the checkpoint kinase Rad53 [211], but if such mechanism exists in metazoans remains to be studied. There are several factors that are involved in the cell cycle regulation of histone gene expression, as will be briefly described below.

There are no obvious sequence elements in RD histone gene promoters to explain the transcriptional activation. However, there are so-called Subtype-Specific Consensus Sequences (SSCSs) in the promoters of particular RD histone genes. SSCSs are bound by specific *trans*-acting factors that regulate transcription, *e.g.* Octamer-binding Transcription Factor 1 (OTF-1) binds H2B promoters and stimulates transcription *in vitro* [212]. Furthermore, in H2A, H3 and H4 genes there is a sequence element called Coding Region Activating Sequence (CRAS), necessary for their expression [213, 214]. And H4 genes have a shared promoter element required for high levels of transcription that is recognised by HiNF-P [215]. The best known transcription factor necessary for the G1/S-phase transition is NPAT [216], which was shown to also depend on SSCSs [217]. In late G1, Cyclin E/Cdk2 is activated and phosphorylates NPAT, which in turn activates transcription of RD histone genes [203, 216, 218]. As mentioned above, NPAT is an essential component of the HLB throughout the cell cycle. It is a large (1427 aa) protein that forms a scafffold for the HLB [196, 219]. It interacts with FLASH to recruit it to the HLB, its phosphorylated form specifically marking S-phase HLBs [218]. During S-phase, many factors involved in RD histone gene expression, *e.g.* Symplekin and Pol II, are recruited to HLBs and thus positioned into close proximity to the genes [197]. Holo-U7 snRNP may form only in S-phase. The U7 snRNP is a permanent component of the HLB, which also undergoes important changes when the cell goes into S-phase. Its protein component Lsm11 binds FLASH which in turn recruits the HCC to then catalyse the 3'-end processing reaction [187, 193, 220].

SLBP plays a major role in enhancing the 3'-end processing reaction of RD histone pre-mRNA in S-phase. SLBP is cell cycle regulated, mainly through translational activation

as cells progress from G1 to S-phase [221]. SLBP levels are maintained high throughout the S-phase and once DNA replication is concluded at the end of S-phase, SLBP is phosphorylated and targeted for degradation [222,223]. This regulation leads to high SLBP protein levels specifically in S-phase, allowing SLBP binding to the SL of nascent RD histone pre-mRNA and enabling 3'-end processing in coordination with DNA replication. Besides its function in 3'-end processing, SLBP plays crucial roles in downstream events, *e.g.* being essential for efficient translation of RD histone mRNAs [180–182, 224] and protecting the mRNA from 3'-5' degradation [167, 225].

At the end of S-phase or when DNA replication is inhibited in S-phase, RD histone mRNA is rapidly degraded to avoid continued translation and so accumulation of excess histone proteins [226,227]. Active translation is essential for RD histone mRNA degradation at the end of S-phase suggesting it depends on an interaction between the terminating ribosome and SLBP. The distance between the stop codon and the SL is critical for this communication [180,227]. Recent studies show that the signal initiating RD histone mRNA decay is a defect in translation termination: the stalled ribosome at the termination codon recruits Upf1, which interacts with the SLBP/3'hExo complex bound to the SL. Upf-1 is activated by Smg1, however it is unclear if this activation occurs before or after Upf-1 binding [182, 225]. 3'hExo is a 3'-5' exonuclease that only trims 2 nucleotides off the RD histone mRNA 3'-end, further exonucleolysis being blocked by SLBP (as described in Subsection 1.2.2). Opening of the SL is required for further degradation into the SL by 3'hExo, possibly through Upf1, which has helicase activity and is essential for histone mRNA decay [228]. Upf1 also recruits TUT7 that adds oligoU-tails to the mRNA 3'-end [229] that in turn can bind Lsm1-7 which promote decapping [230] and 5'-3' degradation by Xrn1. However, 3'-5' degradation by the exosome plays the major role in RD histone mRNA decay [225, 231]. Therefore, exonucleolysis from both mRNA ends is important for RD histone mRNA decay at the end of S-phase.

### 1.3.3 Translation of replication-dependent histone mRNA

RD histone mRNA export is also mediated by NXF-1, as described for polyadenylated mRNA in Subsection 1.1.2 [232]. Mature cytoplasmic RD histone mRNA is bound by the CBC proteins CBP20 and CBP80 at the 5'-end, and SLBP and 3'hExo at the 3'-end. SLBP

is essential for RD histone mRNA translation and its interaction with SLBP-Interacting Protein 1 (SLIP1) stimulates translation [181]. SLIP1 interacts directly with translation initiation factor eIF3g and the mRNA-export factor DBP5 [233]. Recent reports suggest that the CBC proteins may not be substituted by eIF4E, as it is the case for polyadenylated mRNA (as described in Subsection 1.1.3) [234]. The authors suggest an interaction between SLBP and CBP80/20-dependent translation initiation factor (CTIF) that stabilises the mRNA and blocks the exchange for eIF4E (see Figure 1.6). Given the short half-life of histone mRNA, it is possible that histone mRNA is primarily CBP80/20-dependent [225].

### 1.3.4 Summary illustration: Life cycle of "bulk" polyadenylated mRNA *vs* RD histone mRNA

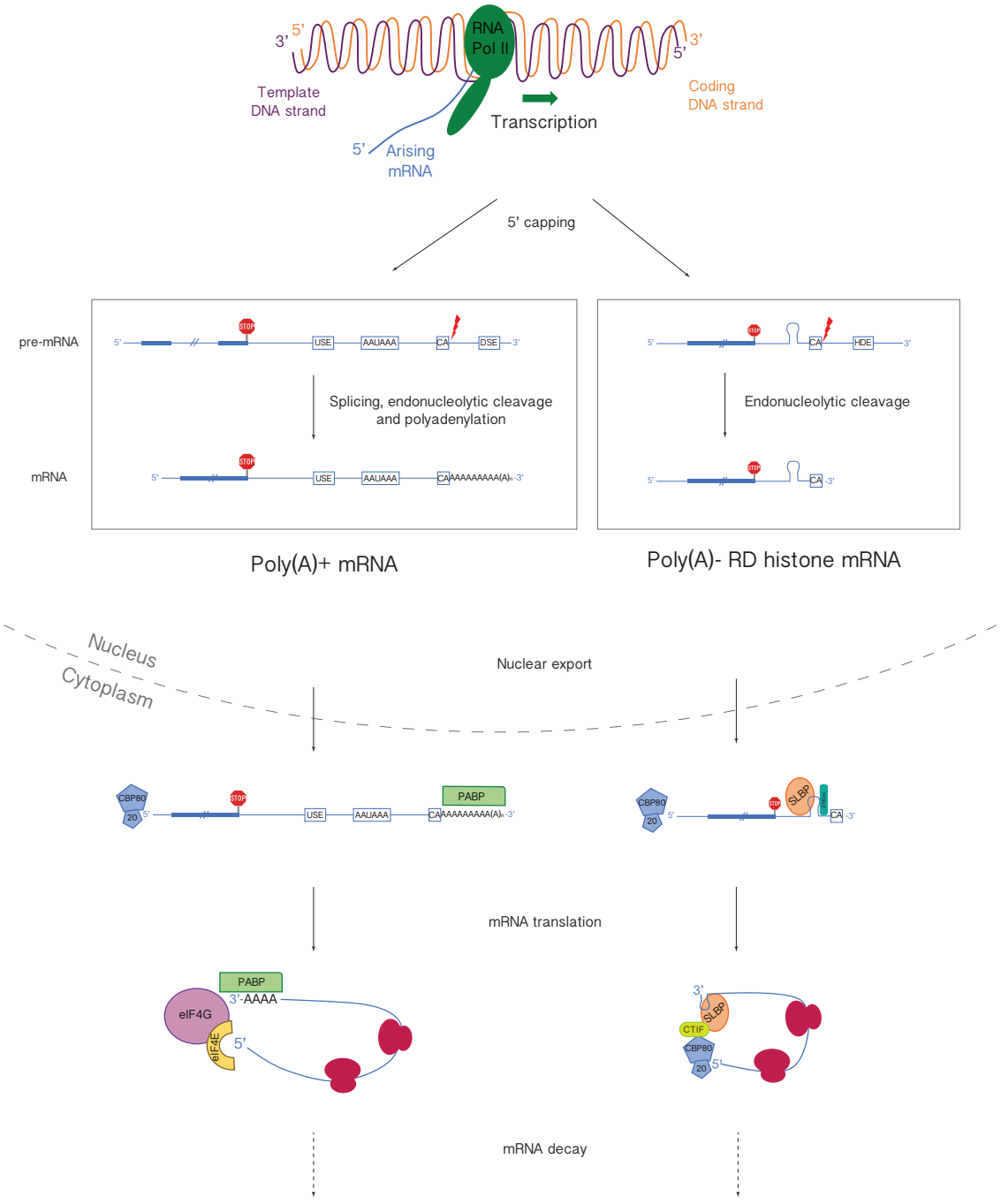



**Figure 1.6 Differences between bulk polyadenylated mRNA and RD histone mRNA.** Pre-mRNA maturation *trans*-acting factors are not depicted in this illustration as they are shown in Figures 1.2 and 1.4.

### 1.3.5 Replication-independent histone genes

Histone proteins are not only essential for DNA compaction into nucleosomes to form chromatin but are also important modulators of gene expression, being continuously postranslationally modified, relocated, removed and redeposited, so changing the chromatin environment [235–237]. Apart from the canonical RD histones mentioned in Section 1.2.2, there are variant histone proteins that differ in protein structure, thus possessing alternative regulatory functions and influencing chromatin accessibility. These variants are called replication-independent (RI) histone proteins. As the name indicates, expression of RI histone genes is not specifically linked to S-phase but occurs throughout the cell cycle. This enables chromatin remodelling and histone replacement also outside of S-phase when canonical histone protein synthesis is inhibited.

Importantly, except from H2AFX, RI histone genes do not contain the SL structure and the mRNA is generated by CPA, thus ending in a poly(A)-tail. Interestingly, many RI genes have introns. In humans there are several RI histone proteins (see Table 1.1). Note that there are no H4 variants.

**Table 1.1** Human histone variants

| Canonical histone | Histone variants |
|---|---|
| H1 | H1.0, H1x, H1T2, H1-oo |
| H2A | macroH2A.1, macroH2A.2, H2A.Bbd, H2A.Z, H2A.X |
| H2B | hTSH2B |
| H3 | CENP-A, cenH3, H3.1t, H3.3 |
| H4 | No known variants |

Many histone variants are expressed at a specific developmental stage or in a tissue-specific manner, such as the H1 variant H1.0 that is mainly expressed in differentiated cells, and H1T2 and H1-oo that are specifically expressed in testes and oocyte, respectively [238]. Other RI histones are ubiquitously expressed, such as CENP-A and cenH3 that are mainly found at centromeres and are important for kinetochore assembly, or H3.3 that replaces the RD histone H3 at actively transcribed genes also having important epigenetic implications [239]. The vertebrate-specific variants macroH2A and H2A.B have important implications in gene silencing or activation, respectively. H2A.Z has roles in transcription initiation, whereas H2A.X is best known for its role in DNA damage signalling [240].

For detailed reviews about the functions of variant histones and associated mechanisms see [235–237, 241]. A detailed description of H2A.X follows in Section 1.5.

## 1.4 Polyadenylation of replication-dependent histone mRNA

RD histone mRNA is known as the only non-polyadenylated mRNA in metazoans. Nevertheless, downstream of the U7-dependent 3'-end processing site of most RD histone genes there are so-called "cryptic polyadenylation sites" and in certain situations RD histone mRNA can be processed at this polyadenylation site (PAS) and acquire a poly(A)-tail.

Experiments involving depletion of factors, directly or indirectly involved in RD histone pre-mRNA 3'-end processing, often lead to the production of some polyadenylated RD histone mRNA. This is caused by inefficient RNA cleavage at the U7-dependent processing site and Pol II transcribing further downstream until it reaches a PAS that is then used for 3'-end processing by CPA. Some examples of factors that lead to this phenotype after RNA interference (RNAi) or mutation are: SLBP [242] and U7 snRNP [243] in *Drosophila* and mammalian cells, Threonine 4 CTD mutation in chicken cells [244], Ars2 [245], CDK9 [246] and Parafibromin [247] in mammalian cells. These studies led to the conclusion that the use of cryptic PASs is widespread in vertebrates. Given the short intergenic distances between RD histone genes, the use of cryptic PAS has been suggested to be a fail-safe mechanism to avoid Pol II "read-through" into a close-by histone gene, which could have disastrous consequences for the cell [248, 249].

The production of polyadenylated RD histone mRNA in unaltered cells has been less studied. Two studies show the accumulation of such mRNA in differentiated cells [250, 251]. Differentiated cells only express RI histones but it is possible that they require a certain amount of canonical histones for replacement, particularly in long-lived cells and especially the canonical histone H4, for which no variants exist [251]. RD histone genes are unable to express the SL mRNA form outside of S-phase, thus in differentiated cells transcribed RD histone genes can only generate polyadenylated mRNA, processed at the polyadenylation site, by CPA.

The sole example of a histone gene that has both replication-dependent and replication-independent characteristics is the H2AFX gene, encoding for the H2A.X histone. This

gene expresses two different mRNA isoforms in cycling mammalian cells. Notably this gene has been intensely studied in the DNA damage field as phosphorylated H2A.X (referred to as $\gamma$H2A.X) is a well recognised marker for DNA damage repair. This gene always gives rise to a S-phase-specific RD histone mRNA processed by the U7-dependent 3'-end mechanism, and an extended polyadenylated mRNA resulting from CPA that is expressed outside of S-phase. Both encode for the same protein [252]. See Section 1.5 for a detailed description of this gene.

## 1.5 H2AFX

### 1.5.1 H2AFX gene

As described in Section 1.3.1, RD histone genes are physically linked in the genome, forming clusters on chromosome 6 and 1 [195]. However, the human H2AFX gene that encodes the H2A.X protein is located on chromosome 11 (11q23.2-q23.3) [253], independent of the main histone gene clusters. The same is true for the other H2A RI variant genes, all of which are located on different chromosomes [254].

The H2AFX promoter has the regular features of protein coding genes. It is localised ~120 bp upstream of the TSS and includes a TATA-box and two CCAAT sequence elements [255]. Transcription reporter assays showed that the proximal CCAAT element is required for efficient H2AFX transcription [255]. Comparing this element to the CCAAT elements in a canonical RD H2A promoter and in the promoter of the RI variant H2A.Z in gel shift assays, suggested that it directs the binding of a specific transcription factor to the RI H2AFX and H2AFZ promoters but not to canonical H2A promoters [255]. This difference in binding specificity seemed to be directed by short homologous sequences located 10 bp upstream and downstream of the CCAAT element in H2AFX and H2AFZ that are absent in H2A promoters [255]. These results suggested that H2AFX transcription is regulated independently of canonical H2A genes.

### 1.5.2 H2A.X mRNA

H2AFX gives rise to two mRNA isoforms: a short (574 bp) non-polyadenylated mRNA ending in the SL structure typical of RD histone mRNA (herinafter referred to as H2A.X SL mRNA) and an extended (∼1.6 Kb) polyadenylated mRNA (from here on referred to as H2A.X poly(A)$^+$ mRNA) [252]. The former is generated by the U7-dependent 3'-end processing mechanism and predominates in S-phase, whereas the H2A.X poly(A)$^+$ mRNA is generated by CPA and is the main H2A.X mRNA isoform outside of S-phase (see Figure 1.7) [256]. Interestingly, the ratio of SL *versus* poly(A)$^+$ H2A.X mRNA varies extensively between different human cell lines [256].

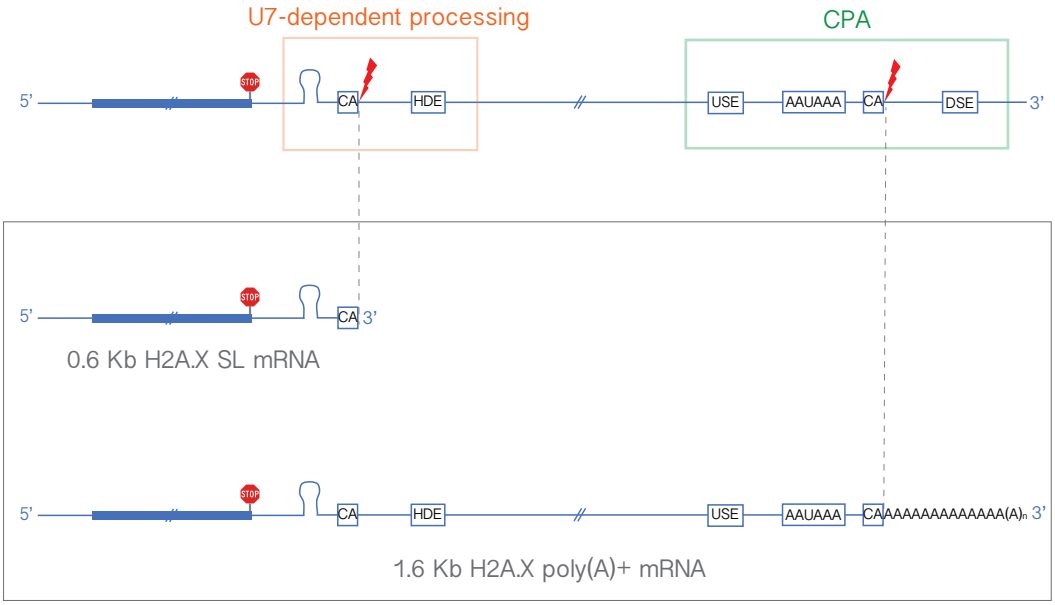



**Figure 1.7 H2A.X mRNA isoforms.** The gray box represents the two mRNA isoforms that can be generated from H2A.X pre-mRNA (shown on top). The H2A.X SL mRNA 3'-end is created by U7-dependent processing and the H2A.X poly(A)$^+$ mRNA 3'-end by CPA. Note that the mature H2A.X poly(A)$^+$ mRNA also contains the SL.

As described in Section 1.3.2, RD histone mRNA levels are tightly coupled to DNA replication, being rapidly downregulated if DNA replication is inhibited to avoid toxic accumulation of histone proteins [226, 227]. Similarly, RD histone mRNA is stabilised if protein synthesis is inhibited, to compensate for histone protein loss [257]. Both these effects are less pronounced for the H2A.X SL mRNA compared to canonical RD H2A mRNA, suggesting H2A.X expression is only weakly linked to DNA replication [256].

Importantly, it was shown that at a posttranscriptional level the H2A.X SL mRNA is regulated in the same way as canonical RD histone mRNA. The distinct regulation of mRNA levels upon inhibition of DNA replication or protein synthesis is only attributable to differences in transcriptional regulation [256]. This in accordance with the differences in the H2AFX and H2A promoters described in Subsection 1.5.1 [255].

Interestingly, the U7-dependent 3'-end processing *cis*-elements (*i.e.* the SL and the HDE) are conserved in all vertebrate but absent in invertebrate H2A.X genes (*e.g. Drosophila*) [223], the latter producing only polyadenylated H2A.X mRNA. This suggests evolutionary pressure to maintain the SL and the H2A.X SL mRNA isoform in vertebrates. Furthermore, SLBP has not only been shown to also bind to the SL of H2A.X poly(A)$^+$ mRNA but it was also suggested to indirectly interact with a highly conserved region in the H2A.X poly(A)$^+$ mRNA, close to the H2AFX polyadenylation site [174].

### 1.5.3 H2A.X protein

H2A.X is present in almost all eukaryotes, from *Giardia lamblia* to humans. The distinctive attribute of H2A.X is the conserved amino acid consensus sequence SQ[E/D]Φ (Serine-Glutamine-Glutamic acid/Aspartic acid and Φ corresponding to a hydrophobic residue). In lower eukaryotes, such as *Giardia lamblia* and *S. cerevisiae*, this motif is encoded in the canonical H2A gene, whereas in higher eukaryotes it is encoded on a separate gene, such as H2AFX in humans, mouse and zebrafish, or H2Av in *Drosophila* [258].

The human H2AFX gene encodes a 142-amino acid (aa) protein, the first 120 aa being nearly identical to the canonical histone H2A, whereas the remaining 22 aa are unrelated [252]. H2A.X is 13 aa longer than canonical H2A and it is this extended C-terminal sequence that contains the conserved SQ[E/D]Φ motif [252]. In mammals, H2A.X comprises ∼2-25% of the total H2A pool [259, 260] and it can substitute both or only one H2A molecule per nucleosome, forming homotypic or heterotypic H2A.X nucleosomes, respectively [254].

The SQ[E/D]Φ motif confers H2A.X an important role in DNA damage. This and other biological functions of H2A.X are discussed below.

### H2A.X in the DNA damage response

The first *in vivo* function for H2A.X was found in 1998, where it was shown that DNA double-stranded breaks (DSBs) lead to phosphorylation of H2A.X at Serine 139 [259]. This residue is embeded in the conserved SQ[E/D]Φ motif in the extended C-terminal tail of H2A.X and the phosphorylated form is called γH2A.X.

Immediately following the formation of a DSB, PI3K-like kinases are activated, including DNA-dependent protein kinase (DNA-PK), ataxia-telangiectasia-mutated (ATM) and ATM/Rad3-related (ATR), that all phosphorylate H2A.X as well as other factors involved in DNA damage signalling [261–264]. γH2A.X is then bound by MDC1 [265], which recruits the MRN complex (MRE11–RAD50–NBS1) [266] that further activates ATM [267], amplifying the signal. γH2A.X specifically forms on nucleosomes at the DSB site and can then extend around it often involving megabase lengths of DNA [268, 269]. This effect has been suggested to act as a platform for the binding of DNA repair factors [270, 271]. Even though γH2A.X is dispensable for the initial recognition of DSBs [272] it has been shown to be required to maintain the stable association of repair factors at DSB sites [272, 273]. This focal concentration of DNA damage repair factors through γH2A.X has been suggested to be important to tether the two broken DNA ends together [274]. Furthermore, after the formation of a DSB, the chromatin around the DSB decondenses to a more accessible conformation, which initially occurs in a γH2A.X-independent manner but that is maintained through γH2A.X [273–275]. This is also supported by the fact that γH2A.X dephosphorylation attenuates DNA damage signalling and is required for DNA damage checkpoint recovery, as persistent γH2A.X phosphorylation leads to G2/M cell cycle arrest [273, 276–278].

DSBs can be caused by multiple endogenous and exogenous factors [279]: direct attack by DNA damaging agents (*e.g.* ionising radiation [259]), reactive oxygen species (ROS) [280], metabolic processes (*e.g.* apoptosis [281] and replication stress [282]), programmed biological processes (*e.g.* class-switch recombination [283]), telomere erosion [284], deficiency of other repair mechanisms (*e.g.* nucleotide-excision repair [285]). Even though γH2A.X is recruited to all these DSBs and is an excellent marker for DSBs, it is not clear how important γH2A.X is in these processes. An example of a mechanism where γH2A.X forms

at DSBs but is dispensable is V(D)J recombination [286, 287]. This raised the hypothesis that γH2A.X is required for the repair of DSBs in which no other molecular recombination program is involved in holding together the broken DNA ends, *e.g* the DSBs generated by ionising radiation, whereas it is dispensable in programmed recombination reactions where other factors tether the broken DNA ends (*e.g.* V(D)J recombination). This hypothesis was supported by showing that γH2A.X is necessary for retroviral DNA integration but not for postintegration where DNA ends are anchored covalently by viral DNA [288].

Importantly, not only DSBs lead to γH2A.X formation. γH2A.X also forms at single-stranded DNA breaks induced by replication stress or ultraviolet C irradiation [289, 290]. In this case, γH2A.X does not appear as foci but as diffused pattern distributed throughout the nucleus. This effect is called pan-nuclear signal and has been shown to precede irreversible cell death [290].

At a systemic level, it has been shown that mice lacking H2A.X are more sensitive to DNA damage, show chromosomal abnormalities and are more predisposed to develop cancer [271, 291]. Similarly, in humans the chromosomal region containing H2AFX (11q23.2-q23.3) is often deleted or mutated in a wide range of cancers [279, 291, 292].

Apart from phosphorylation at Serine 139, H2A.X has been shown to contain other residues subject to posttranslational modifications implicated in the DNA damage response. One example is phosphorylated tyrosine 142 (Y142), which was shown to prevent recruitment of repair factors to serine 139 phosphorylation (γH2A.X) instead recruiting pro-apoptotic factors [293]. However, the importance of the Y142 phosphorylated H2A.X form is unclear as in some cell lines it is bearly detectable [260]. Other H2A.X postranslational modifications are discussed elsewhere [294–296].

### Alternative H2A.X functions

Apart from its role in DNA damage signalling, H2A.X is involved in many other cellular processes, including sex chromosome inactivation in germ cells, X chromosome inactivation in somatic cells, chromatin regulation in mitosis, asymmetric sister chromosome segregation, embryonic stem-cell development, maintenance of cellular senescence, as well as epithelial-mesenchymal transition [297–299].

The role of H2A.X in meiotic sex chromosome inactivation (MSCI) was indicated by the fact that H2A.X-deficient male but not female mice are infertile and show X-Y chromosome asynapsis [271]. Later it was shown that H2A.X-deficient spermatocytes have impaired sex-body formation, fail to initiate MSCI and show defects in meiotic pairing. These results proved a role for $\gamma$H2A.X in sex chromosome condensation and transcriptional inactivation [300]. Furthermore, $\gamma$H2A.X and BRCA1 were shown to associate with the inactive X-chromosome specifically in late S-phase and are suggested to play a role in decatenation after DNA replication and to re-establish the optimal chromatin environment for the binding of chromatin remodelling complexes [301].

Interestingly, normally proliferating cells show DNA damage-independent fluctuations in $\gamma$H2A.X signal during the cell cycle, reaching maximum levels at M-phase [302], however it is debatable if that is always the case. M-phase $\gamma$H2A.X specifically localises to kinetochores where it recruits proteins involved in spindle assembly checkpoint regulation [303]. Another very interesting paper suggests a role for $\gamma$H2A.X in non-random chromosome segregation [304]. In this scenario, a DSB created on the new DNA strand due to replication stress would lead to $\gamma$H2A.X recruitment and spreading into distant parts over the chromosome, marking the DNA as "damaged". If $\gamma$H2A.X spreading reaches a centromere, it may lead to a change in chromatin conformation, triggering kinetochore disfunction. This is recognised by proteins of the spindle assembly checkpoint to enforce cell cycle arrest and thus prevent the segregation of damaged chromosomes [305].

H2A.X also plays a role in development. Thus it was shown to be highly expressed in undifferentiated mouse embryonic stem cells (mESCs) and preimplantation embryos [306]. H2A.X deletion in mESCs and mouse induced pluripotent stem cells (miPSCs) showed reduced self-renewal capacity [307] due to transcriptional activation of genes involved in extraembryonic differentiation [308]. Overall these results show that H2A.X is an important epigenetic regulator of chromatin architecture during development.

$\gamma$H2A.X has also been suggested to play a role in cellular senescence, where persistent DNA damage foci (PDDF) positive for $\gamma$H2A.X lead to p53-dependent growth arrest and p53-independent senescence-associated inflammatory cytokine secretion. This effect was suppressed by H2A.X depletion [297, 309, 310].

Recent work implicates H2A.X in the epithelial-mesenchymal transition [298, 299]. Here

it was shown that H2A.X "knock out" HCT116 cells undergo epithelial-mesenchymal transition due to depletion of H2A.X from promoters of genes involved in this process, subsequently leading to their activation. This shows that H2A.X has important roles in the control of gene expression.

Taken together, both H2A.X and $\gamma$H2A.X have important DNA-damage-independent functions and DNA-damage-related functions that go beyond DSB repair. The unique structure of H2A.X is able to alter the chromatin environment, ultimately influencing gene expression and epigenetic processes.

## 1.6 Aim of this thesis

RD histone mRNA is known as the only non-polyadenylated mRNA in metazoans. However, as described in Section 1.4, certain circumstances can lead to the formation of some polyadenylated RD histone mRNA. This phenotype has mostly been observed in conditions where U7-dependent processing was made inefficient by depleting required factors, leading to polymerase read-through and subsequent 3'-end processing at a downstream cryptic PAS. At the starting point of this thesis very few publications indicated the formation of polyadenylated RD histone mRNA under physiological conditons [250]. Therefore, the primary aim of my thesis was to determine which RD histone genes are able to form correctly processed polyadenylated mRNA that is stable and translated in unaltered conditions (*i.e.* intact U7-dependent processing), ultimately establishing a function for polyadenylated histone mRNA and the protein it encodes. Thus the first experiment performed for this project was an RNAseq analysis that allowed identification of genes that produce polyadenylated RD histone mRNA, which PAS was used, as well as their nucleo-cytoplasmic mRNA distribution. Contemporaneously with my data analysis, an excellent publication demonstrated the accumulation of a subset of polyadenylated RD histone mRNA as part of the terminal differentiation program in all tissues [251]. This together with the observation that H2AFX is the clearest example of a histone gene that produces both poly(A)$^-$ and poly(A)$^+$ mRNA [252] but whose functions had not been described, led me to redirect my research to study H2AFX gene expression.

The H2AFX gene and its mRNA isoforms were first described in 1989 [252]. It was

shown that H2A.X expression is only weakly linked to DNA replication due to differences in the H2AFX promoter compared to RD histone gene promoters [255,256]. Since then, H2A.X is frequently considered a replication-independent histone variant [256]. However, the cell cycle regulation and the function of these mRNA isoforms remains poorly characterised. Furthermore, the fact that all vertebrates have evolved to possess the H2A.X SL mRNA suggests a beneficial function for this mRNA isoform.

Appart from three studies involving cell-line-specific miRNA targeting of the extended 3'-UTR of the H2A.X poly(A)$^+$ mRNA [311–313], H2A.X depletion studies have invariably targeted total H2A.X, without differentiating between the two H2A.X mRNA isoforms. Similarly, is unknown whether the translation of the two H2A.X mRNA isoforms is regulated. It has however been shown that SLBP binds to the SL of the H2A.X poly(A)$^+$ mRNA [174], making it possible that both mRNA isoforms are translated by the SLBP-CBC-dependent mechanism described in Section 1.1.3. Furthermore, it is unclear which role the two mRNA isoforms play in the DNA damage response. The aim of my thesis is therefore to characterise these two H2A.X mRNA isoforms during the cell cycle and in DNA damage signalling.

# Chapter 2

# Materials and Methods

## 2.1 Cell culture

### 2.1.1 Cell lines

The cell lines employed in this study are listed in Table 2.1.

**Table 2.1** Cell lines employed in this study. The HeLa Flp-In T-REx cell line was kindly provided by Matthias Gromeier and Elena Dobrikova at Duke University Medical Center (North Carolina, USA). hTERT: human telomerase reverse transcriptase.

| Cell line | Organism | Cell type | Tissue | Disease | Obtained from |
|---|---|---|---|---|---|
| *HeLa* | Human | Epithelial | Cervix | Adenocarcinoma | Proudfoot Laboratory, Oxford, UK |
| *HeLa Flp-In T-REx* | Human | Epithelial | Cervix | Adenocarcinoma | Duke University Medical Center, USA |
| *HCT116* | Human | Epithelial | Colon | Colorectal carcinoma | Marzluff Laboratory, North Carolina, USA |
| *RPE-1* | Human | Epithelial | Retina | Immortalized with hTERT | Marzluff Laboratory, North Carolina, USA |

### 2.1.2 Culturing conditions

HeLa and RPE-1 cells were cultured in Dulbecco's Modified Eagle's Medium (DMEM, Sigma, #D6429) supplemented with 10% fetal bovine serum (FBS, Gibco, #10500-064) and Penicillin-Streptomycin (Gibco, #15140-122). HCT116 cells were maintained in McCoy's 5A media (Sigma, #M8408) supplemented with 10% FBS, Penicillin-Streptomycin and GlutaMAX (Gibco, #35050-061). All the cells were grown in incubators at 37°C and 5% $CO_2$ and periodically examined under the microscope (Carl Zeiss, #491206-0007-000).

### 2.1.3 Cell passaging

To minimise secondary effects due to cell passage number or genetic variability between the same cell line from different laboratories or even from the same laboratory, each cell line employed in this study was expanded immediately upon acquisition and multiple vials were kept at -80°C (see Subsection 2.1.4). Each cell batch was kept for ten passages at most, after which a new vial was thawed.

The cells were passaged after reaching 90% confluency: they were washed with PBS (Sigma, #D8537) and incubated with 1x Trypsin-EDTA (TrypLE Express, Gibco, #12604-013) for 5 min at 37°C. The trypsin was inactivated with culture media (see Subsection 2.1.2), the cells were separated into single cells by pipetting and transferred to a new cell culture dish diluting accordingly (1/10 dilutions for maintenance or less depending on the experiment).

### 2.1.4 Cell freezing and thawing

All four cell lines were frozen in FBS containing 10% Dimethyl sulfoxide (DMSO, Sigma, #D2650). The cells were harvested using trypsin (see Subsection 2.1.3), the trypsin was inactivated using cell culture media, then the cells were transferred to a 15 mL tube (Sarstedt, #62.554.502), spun down for 3 min at 300 x $g$ and the resulting cell pellet was resuspended in 0.9 mL FBS per vial and transferred to 2 mL cryovials (Starlab, #E3090-6222) containing 100 µL DMSO. The cell suspension was mixed several times by inversion and the tubes were transferred to a freezing container (Mr. Frosty, Thermo Fisher Scientific #5100-0001) and deposited at -80°C. The next day the cells were transferred to a -80°C storage box.

To thaw the cells, a vial was taken from the -80°C and incubated at 37°C to allow partial thawing of the cell suspension (usually less than 5 min). As quickly as possible, the cells were transferred to a cell culture dish with the corresponding culturing media and transferred to the 37°C incubator.

### 2.1.5 Cell counting

Cell counting was performed using a haemocytometer (Neubauer Improved, Weber England, #BS.748) or a NC-3000 Advanced Image Cytometer (Chemometec). For the former, 10 µL cell suspension was resuspended in 30 µL Trypan Blue Solution (Sigma, #T8154) to a 1/4 dilution and the cells were counted and their concentration calculated as follows:

$$\frac{\text{Sum of cells counted in the 4 main squares of the chamber}}{100} = \times 10^6 \text{cells/mL}$$

The NC-3000 Advanced Image Cytometer was employed with Solution 13 $AO \cdot DAPI$ (Chemometec, #910-3013) and NC-Slides A8 (Chemometec, #942-0003). 1 µL of Solution 13 was mixed with 19 µL of the single cell suspension. After thorough mixing, 10 µL of the mix was applied to the chamber and analysed.

## 2.2 Cell cycle synchronisation

### 2.2.1 Nocodazole block

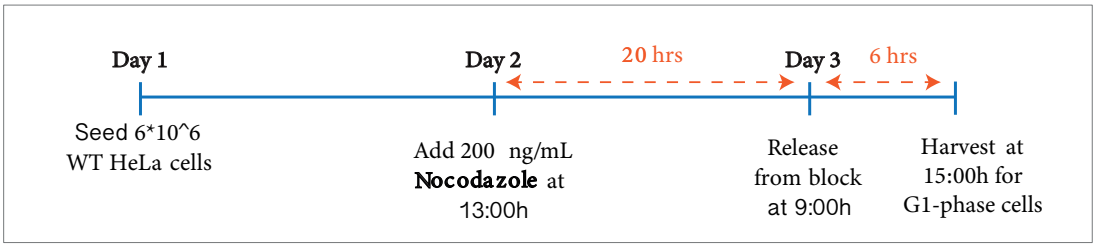



**Figure 2.1 G1-phase synchronisation protocol for HeLa cells, using nocodazole.**

Nocodazole is a microtubule de-polymerisation agent, blocking the cells in mitosis [314]. $6 \times 10^6$ HeLa cells were seeded per 150 mm dish 24 hr before treatment. Nocodazole was added to a final concentration of 200 ng/mL for 18-20 hrs. After that time most cells were in mitosis and thus no longer attached to the cell culture dish. To release the cells from the nocodazole block the floating cells were collected, washed twice with PBS and resuspended in fresh cell culture media. After 6 hr growth, about 90% of the cells were in

G1-phase, which was always verified by flow cytometry. See Figure 2.1 for a schematic example of this protocol.

### 2.2.2 Aphidicolin block

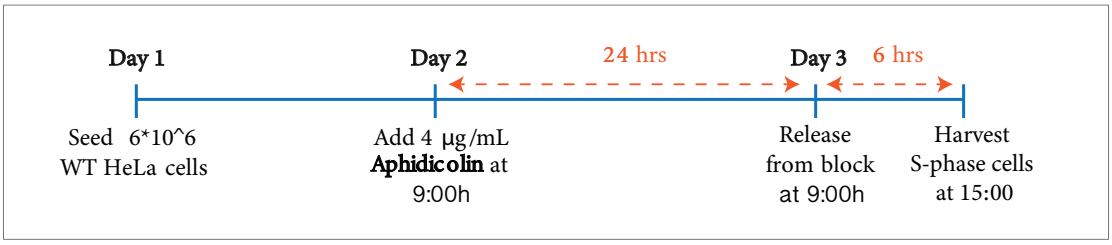



**Figure 2.2 S-phase synchronisation protocol for HeLa cells, using aphidicolin.**

Aphidicolin reversibly inhibits B-type DNA polymerases thus preventing DNA synthesis and blocking cells at the G1/S-phase border [315,316]. 24 hr before the treatment, $6 \times 10^6$ HeLa cells were seeded per 150 mm dish. The following day, 4 µg/mL Aphidicolin (Santa Cruz, #38966-21-1) were added to the cell culture media and incubated for 24 hrs. To release from arrest and allow S-phase entry, the cells were washed three times with PBS, collected by trypsinisation (see Subsection 2.1.3), spun at 300 x *g*, washed with PBS one more time, spun at 300 x *g* and resuspended in fresh culture media. After 6 hr growth, close to 90% of cells were in S-phase (verified by flow cytometry). A schematic example of this protocol is shown in Figure 2.2.

### 2.2.3 Double thymidine block

An excess of thymidine reversibly inhibits dCTP synthesis thus causing a dNTP imbalance and so blocking cells at the G1/S-phase border [317]. To synchronise HeLa cells, $1 \times 10^6$ cells were seeded per 10 cm dish, in as many dishes as time points collected. The following day, thymidine was added to a final concentration of 2 mM and incubated for 19 hrs. The cells were released from the first thymidine block by washing them twice with PBS followed by the addition of fresh culture media. After 9 hrs, 2 mM thymidine was added and incubated for 18 hr to arrest the cells for a second time. After the release from the second block time points were collected. The cells from each 10 cm dish were split in three to obtain a sample for flow cytometry, protein and RNA (10% for flow cytometry, 25% for

protein and 65% for RNA). See Figure 2.3A for an example of the HeLa double thymidine block (DTB) protocol. The DTB protocol for HCT116 cells is very similar, only differing in the number of cells seeded for the experiment and in the thymidine incubation periods (see Figure 2.3B).

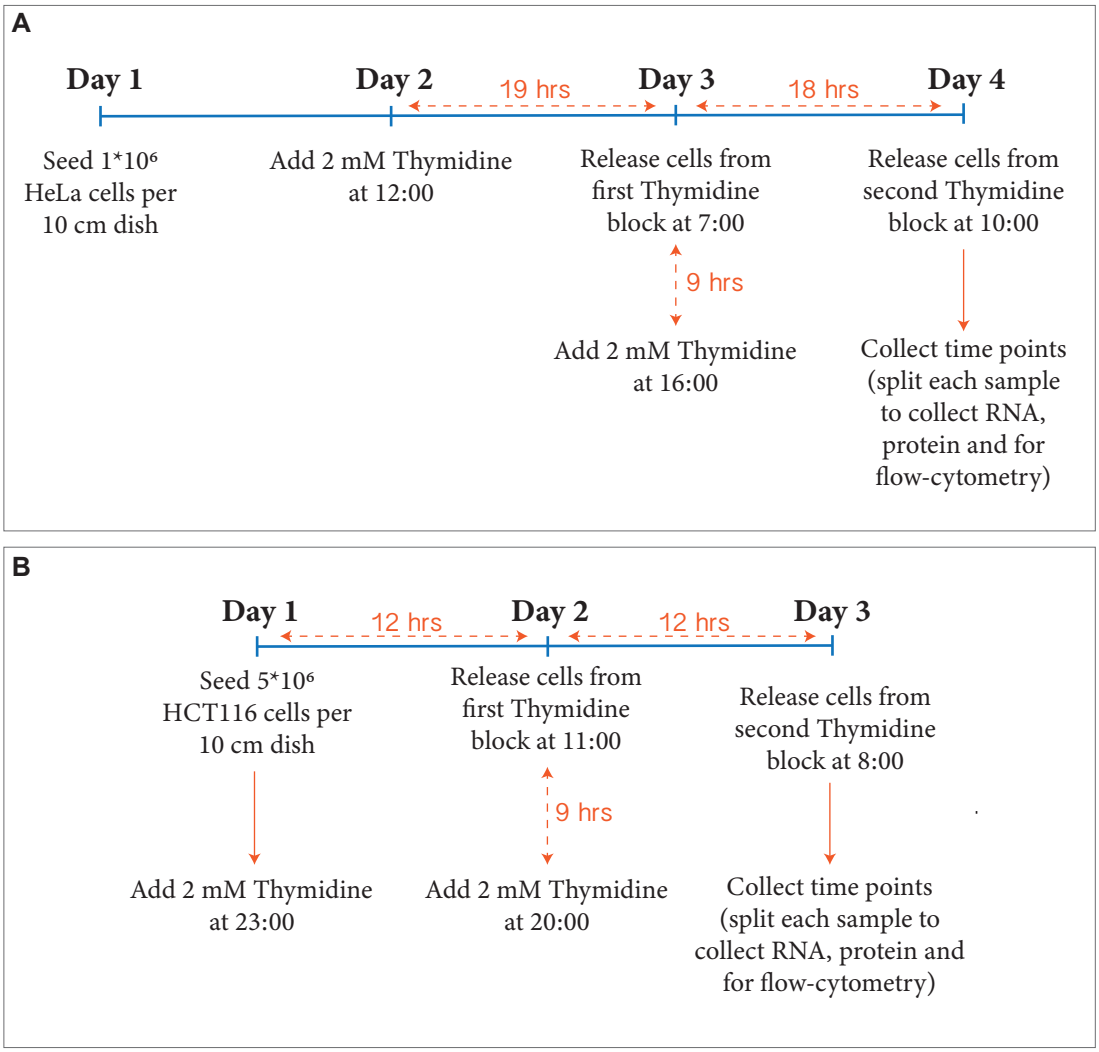



**Figure 2.3 Double thymidine block for cell cycle synchronisation of HeLa cells (A) and HCT116 cells (B).**

### 2.2.4 RPE-1 cell cycle arrest by contact inhibition

RPE-1 cells were seeded to be subconfluent 24 hr later. For RNAi experiments, the cells were transfected at the subconfluent state. The day after the transfection, the cells were 100% confluent and then incubated for a total of 4 days. During this time the cells undergo contact inhibition entering $G_0$ arrest. The cells were then split 1/10 into a new cell culture

dish to allow cell cycle re-entry and harvested at the desired time point over a period of about 22h.

## 2.3 Flow cytometry

The initial flow cytometry experiments were performed using propidium iodide staining. Later, Jean Cook (Chapel Hill, North Carolina) very kindly provided me with the EdU staining protocol so that all the following experiments were carried out using EdU (5-Ethynyl-2'-deoxyuridine) and DAPI (4', 6-diamidino-2-phenylindole) staining.

### 2.3.1 Propidium iodide staining

The cells were harvested using trypsin and thoroughly resuspended to disrupt any cell clumps. After pelleting the cells at 300 x *g* at room temperature, they were washed once with PBS and repelleted. The supernatant was removed and the cell pellet was resuspended in 50 µL PBS. For fixation, the tube containing the cell suspension was placed on a vortex and 1 mL 70% ethanol was added dropwise to avoid clumping. At this point, samples were stored at 4°C or directly processed further. Before staining, cells were washed once with PBS and then resuspended in up to 0.5 mL staining solution (PBS with 0.1% Triton X-100 (Sigma, #T8787), 0.2 mg/mL DNAse-free RNAse A (Thermo Fisher Scientific, #10753721) and 0.02 mg/mL propidium iodide (Sigma, #P4170). After sample incubation for 30 min at room temperature in the dark, the samples were analysed by flow cytometry.

### 2.3.2 EdU labelling

The cells were treated with 10 µM EdU (Abcam, #ab146186) for 30 min, allowing EdU incorporation into the newly synthesised DNA. As described above, the cells were harvested using trypsin, washed once with PBS and throrougly resuspended in 50 µL PBS. In this case, the cells were fixed with 0.5 mL 4% formaldehyde (Sigma, #F1635) in PBS for 15 min at room temperature. To help pellet the cells after fixation, 1 mL 1% BSA in PBS was added. At this stage, samples could be stored at 4°C or directly processed further. Cells were pelleted by spinning at 1000 x *g* at room temperature for 3-5 min. All spins in this protocol were performed using these conditions. The supernatant was discarded and

cells were resuspended in 1 mL PBS with 1% BSA and 0.5% Triton X-100 and incubated at room temperature for 15 min for permeabilisation. For the staining, the cells were pelleted and resuspended in 0.5 mL labeling reaction: 1 mM $CuSO_4$ (Copper II sulfate pentahydrate, Sigma, #C8027), 0.2-2 µM Alexa Fluor 647 Azide (Triethylamoniumsalt, Thermo Fisher, #A10277), 100 mM L-ascorbic acid (Sigma, #A7506) in PBS. Importantly, the L-ascorbic acid was prepared fresh every time and only added to the labelling reaction mix immediately before use. The cells were stained for 30 min in the dark and 1 mL of PBS with 1% BSA and 0.5% Triton X-100 was added before pelleting the cells. The cell pellet was resuspended in DAPI staining solution: 1 µg/mL DAPI (Thermo Fisher Scientific, #62248), 100 µg/mL DNAse-free RNAse A (Thermo Fisher Scientific, #10753721) in PBS with 1% BSA and 0.5% Triton X-100. After incubating over night at 4°C or for 1 hr at 37°C the samples were analysed by flow cytometry.

### 2.3.3 Flow cytometer

The samples stained with propidium iodide were analysed with the BD FACS Calibur and EdU-labelled cells with BD LSRFortessa X-20. In cases where the latter could not be used, the Cytek DxP8 was employed.

#### BD FACS Calibur

BD FACS Calibur was employed to analyse cells stained with propidium iodide. Propidium iodide is detected with the 488 nm laser using a 585/42 band pass filter.

#### BD LSRFortessa X-20

Samples labelled with EdU (Alexa 647) and stained with DAPI were analysed with BD LSRFortessa X-20. DAPI was detected with the 405 nm laser, a 450/50 band pass filter and a 420 nm long pass filter. Alexa 647 was detected using a 640 nm laser, a 670/30 band pass filter and a 665 nm long pass filter.

#### Cytek DxP8

To detect DAPI with the Cytek DxP8, the 405 nm laser was employed with a 445/50 band pass filter, whereas Alexa 647 was detected using a 640 nm laser and a 661/16 band pass filter.

## 2.4 Transient transfection

### 2.4.1 siRNA transfection

HeLa and HCT116 cells were transfected with siRNAs using Lipofectamine RNAiMAX (Invitrogen, #13778-150) and RPE cells using Dharmafect 1 Transfection Reagent (Dharmacon, #T-2001-01), according to the manufacturers's instructions. See Table 2.2 for siRNA sequences or catalog numbers.

**Table 2.2** siRNAs employed in this study.

| siRNA | Company | Sequence or supplier ID |
|---|---|---|
| siLuc (custom siRNA) | Sigma | Sense sequence (5'-3') : GAUUAUGUCCGGUUAUGUAUU<br>Antisense sequence (5'-3'): [phos]UACAUAACCGGACAUAAUCUU |
| siRNA pool (ON-TARGETplus Human H2AFX siRNA) | Dharmacon | L-011682-00-0005 |
| Single siRNA (custom human H2AFX siRNA) | Integrated DNA Technologies | Sense sequence (5'-3') : CAAUCCAAGCACCUAGAUACCAGCA<br>Antisense sequence (5'-3'): UGCUGGUAUCUAGGUGCUUGGAUUGCC |

### 2.4.2 Plasmid transfection

Plasmids were transfected using Lipofectamine 2000 (Invitrogen, #11668-019), according to the manufacturer's instructions.

## 2.5 Cloning

### 2.5.1 Software

The software employed for cloning was SnapGene, version 4.2.11.

### 2.5.2 Transformation

50 µL competent XL1-Blue *E. coli* cells were thawed on ice and added to a 1.7 mL tube (Axygen, Corning, #MCT-175-C) containing the plasmid DNA to be transformed (10 ng purified plasmid DNA or 5 µL of a ligation reaction or a DpnI-treated site-directed mutagenesis PCR reaction). After incubating on ice for 20 min, the bacteria were subjected to a 45 sec heat-shock at 42°C. After incubating on ice for 2 min, 250 µL Luria-Bertrani broth (LB, 1% w/v Tryptone (Oxoid, #LP0043), 1% w/v NaCl (VWR, #27810.295) and 0.5% w/v yeast extract (Oxoid, #LP0021), adjusted to pH7 with 5 N sodium hydroxide (VWR, #28244.262) and sterilised by autoclaving) were added and the bacteria were incubated on a Thermomixer (Eppendorf, #05-400-203) at 37°C, shaking at maximum speed for 1 hr. For ligations, the bacteria were spun at 6800 x *g* for 3 min, the supernatant was discarded leaving about 50 µL to resuspend the bacteria and everything was employed for plating. For purified plasmids or site-directed mutagenesis, 10 and 100 µL of the bacteria suspension were plated. Agar plates (1.5% w/v Agar (Oxoid, #LP0011), 1% w/v Tryptone T (Oxoid, #LP0043), 1% w/v NaCl and 0.5% w/v yeast extract (Oxoid, #LP0021), autoclaved, cooled and supplemented with antibiotics) supplemented with the corresponding antibiotics (100 µg/mL ampicillin (Melford Biolaboratories, #A0104)) were incubated over night at 37°C. Colonies were either directly employed for colony PCR (see Subsection 2.5.3) or picked and transferred to 5 mL ("mini culture") LB supplemented with ampicillin followed by incubating about 16 hr at 37°C.

### 2.5.3 Colony PCR

For colony PCR, 1 µL of a "mini culture" or a picked colony was added to 20 µL PCR reaction: 0.2 µL Phusion polymerase (New England Biolabs, #M0530S), 4 µL 5x GC buffer (New England Biolabs, #M0530S), 0.4 µL 10 mM dNTPs (New England Biolabs, #N0447S), 1 µL of the 10 µM dilution of each primer, 2 µL 5 M Betaine (Sigma, #61962) and 11.4 µL nuclease-free water (Ambion, #AM9937). Importantly, for colony PCR the initial denaturation step was carried out for 3 min to disrupt bacteria. The melting temperature was determined using the NEB Tm calculator (`https://tmcalculator.neb.com/#!/main`) and the remaining parameters were the suggested in the Phusion polymerase protocol

(New England Biolabs, #M0530S).

### 2.5.4 H2AFX 3'UTR reporter cloning for transient transfection

A plasmid containing the EGFP ORF under the control of a HIST2H4 promoter (pHIST2H4-EGFP, see top plasmid map in Figure 2.4), kindly provided by Marc Ruepp (University of Berne, Switzerland), was cut with XbaI (New England Biolabs, #R0145S), gel-purified using the Zymoclean Gel DNA Recovery Kit (Zymo Research, #D4007), dephosphorylated using the FastAP Thermosensitive Alkaline Phosphatase (Thermo Fisher Scientific, #10819230) and employed for insert ligation (see below).

Either the entire H2AFX 3'-UTR or only the region until downstream of the HDE were amplified from genomic DNA with primers containing an XbaI restriction site (*H2AX-3UTR-XbaI-Fw* and *H2AX-3UTR-1kb-XbaI-Rev* or *H2AX-3UTR-XbaI-Fw* and *H2AX-3UTR-SL-HDE-XbaI-Rev*, respectively, from Table 2.3). The PCR products were gel-purified using the Zymoclean Gel DNA Recovery Kit (Zymo Research, #D4007) and employed for ligation into dephosphorylated, XbaI-linearised plasmid described above. The ligated constructs were transformed in XL1-Blue competent *E. coli* bacteria as described in Subsection 2.5.2 and plated on ampicillin-resistant agar plates.

Positive clones were selected using colony PCR, as described in Subsection 2.5.3. For the construct containing the entire H2A.X 3'-UTR, primers *EmGFP-Fw* and *H2AFX-polyA-Rev* were employed and for the construct containing the short H2A.X 3'-UTR, primers *EmGFP-Fw* and *bGH rev* were employed. Positive clones were grown in LB supplemented with 100 µg/mL ampicillin and plasmid DNA was isolated as described in Subsection 2.7.3. Sanger Sequencing (2.8.1) was performed with the same primers.

The HDE region from both constructs was deleted using the QuikChange Multi Site-Directed Mutagenesis Kit (Agilent, #200515) and the primers listed in Table 2.3. Positive clone selection was performed by Sanger Sequencing as described above.




**Figure 2.4 Plasmid maps of H2AFX 3'-UTR reporter.** The plasmid on the top shows the original plasmid containing the HIST2H4 promoter and the EGFP ORF. The plasmid DNA sequence was obtained from Marc Ruepp (University of Berne, Switzerland) and employed to create the plasmid map using the SnapGene software. The two plasmid maps on the left show the resulting plasmid after inserting the entire H2AFX 3'-UTR (top left) and the short H2AFX 3'-UTR ending downstream of the HDE (lower left). The two plasmid maps on the right, show the HDE deletion mutants from the plasmids on the left.

**Table 2.3** Primer sequences for PCR amplification of the insert from human genomic DNA, primers employed to delete the H2AFX HDE by site-directed mutagenesis and primers employed for colony PCR and Sanger Sequencing.

| **Primer name** | **Sequence (5'-3')** |
|---|---|
| H2AX-3UTR-XbaI-Fw | atcagcTCTAGAgagggcccgcgccgcggccg |
| H2AX-3UTR-1kb-XbaI-Rev | tagtcgTCTAGAtcctatggtttgccccttcag |
| H2AX-3UTR-SL-HDE-XbaI-Rev | tagtcgTCTAGActtgccccgcagtctgaagc |
| **Site-directed mutagenesis primers** | |
| HDE-mut-Quik-Fw | ccaccgccctcatggagccgcttcagac |
| HDE-mut-Quik-Rev | gtctgaagcggctccatgagggcggtgg |
| **Colony PCR and sequencing primers** | |
| EmGFP-Fw | CAAAGACCCCAACGAGAAGC |
| H2AFX-polyA-Rev | CCCAGCGCAGACCTATGAAT |
| bGH-Rev | TAGAAGGCACAGTCGAGG |

### 2.5.5 H2AFX 3'UTR reporter cloning for stable integration into HeLa-FlpIn cells

For the stable integration of H2AFX gene constructs into the HeLa-FlpIn cell line, the pcDNA5-FRT-TO-eGFP-Linker plasmid, kindly provided by the Castello laboratory (Oxford University), was employed as backbone. Furthermore, this plasmid contains the FRT site required for stable integration and the Hygromycin B resistance gene as selection marker (plasmid map shown in Figure 2.5).

For constitutive expression of the H2AFX gene controlled by its own promoter and poly(A) site, the region of the plasmid containing the CMV promoter up to after the bGH (bovine growth hormone) poly(A) site was excised using MluI and SphI restriction enzymes (New England Biolabs, #R0198S and #R0182S, respectively). Note that the Tet-operator was also excised. The linearised plasmid was gel-purified using the Zymoclean Gel DNA Recovery Kit (Zymo Research, #D4007) and dephosphorylated using the FastAP Thermosensitive Alkaline Phosphatase (Thermo Fisher Scientific, #10819230). For the insert, the H2A.X gene was amplified from HeLa genomic DNA using 5'-CTATCGacgcgtTTCCCAGACGCTCTCTAGGT-3' as forward primer, binding 335 base pairs (bp) upstream of the H2AFX 5'UTR, thus including the H2AFX promoter, and containing the MluI restriction site. As reverse primer, 5'-TTACATgcatgcGTTCTCCTGGTAC GTCCTTTCT-3' was employed, binding 1807 bp downstream of the H2AFX 5'UTR start, including the entire H2AFX gene with its own poly(A) site and containing the SphI

restriction site. The PCR was performed with Q5 polymerase (New England Biolabs, #M0491L), using 100 ng HeLa genomic DNA as template, at a melting temperature of 67°C, 1 min and 15 sec elongation, for 35 cycles. The remaining PCR parameters were set according to the manufacturer's instructions. The PCR product was precipitated using 100% ethanol (see Subsection 2.7.1), gel-purified as described above, cut with MluI and SphI, gel purified once more and ligated into the linearised and dephosphorylated plasmid. Ligation was performed using T4 DNA ligase (New England Biolabs, #M0202S). The ligated product was transformed into *E. coli* XL1-Blue competent cells as described in Subsection 2.5.2. Positive clones were selected by colony PCR (see Subsection 2.5.3) using *Seq Primer 1-Fw* and *Seq Primer 1-Rev* from Table 2.4, binding upstream and downstream of the insert, respectively. As a final confirmation, the positive clones were sent for Sanger Sequencing (see Section 2.8.1) using all the primers listed in Table 2.4. The resulting plasmid called pcDNA5-FRT-H2AFX_wt was employed as a template to make H2AFX mutant constructs (see Section 2.5.5).

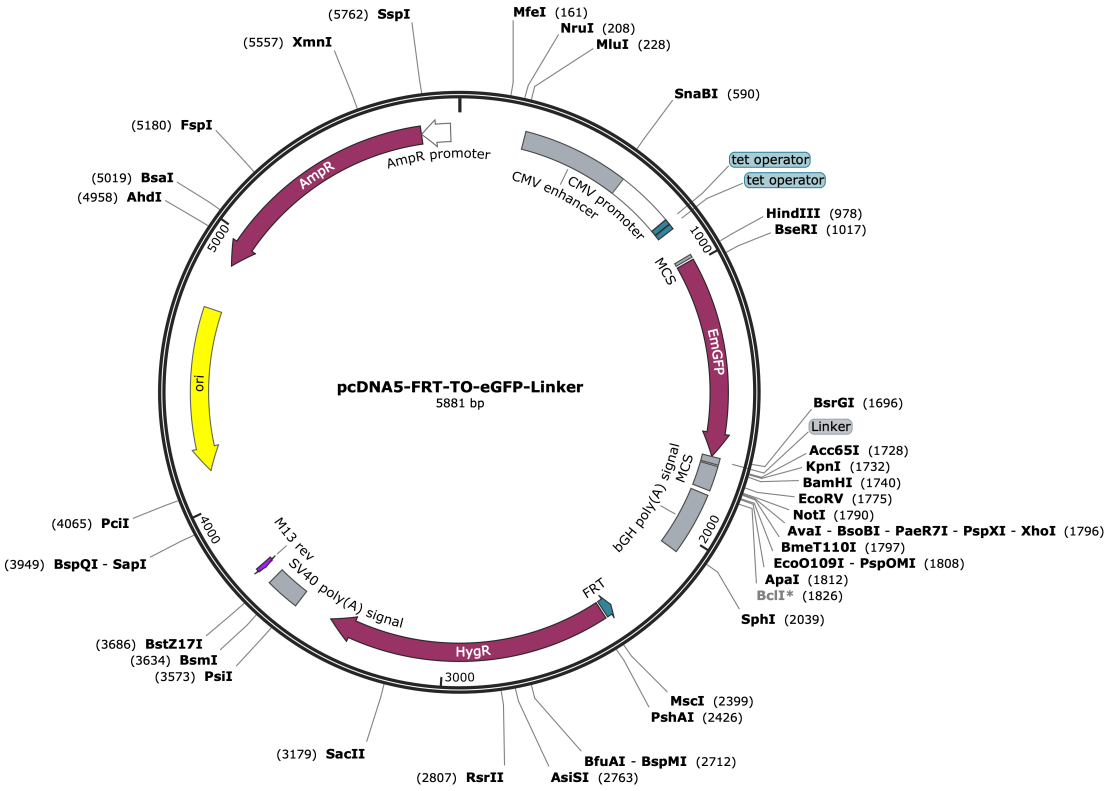



**Figure 2.5 Original plasmid employed for stable integration into the HeLa-FlpIn cell line.** The MluI and SphI restriction sites were employed to clone in the H2AFX gene with its own promoter. The plasmid DNA sequence was obtained from the Castello Laboratory (University of Oxford) and employed to create the plasmid map in the SnapGene software.

**Table 2.4** Primers employed for Sanger Sequencing of pcDNA5-FRT-H2AFX constructs.

| **Primer name** | **5'-3' sequence** |
|---|---|
| Seq Primer 1-Fw | AAGCCAGTATCTGCTCCCTG |
| Seq Primer 2-Fw | CGGGTATTGGAGAAAAGAGCC |
| Seq Primer 3-Fw | CGGCAGTGCTGGAGTACC |
| Seq Primer 4-Fw | TGCTGCCCAAGAAGACCAG |
| Seq Primer 5-Fw | GTAGGGTTCGGGCCTTCC |
| Seq Primer 6-Fw | CAGGCCTTTCACATCAGCTC |
| Seq Primer 7-Fw | CTCAACTCGGCAATCCAAGC |
| Seq Primer 1-Rev | CTTAATGCGCCGCTACAGG |

### Site-directed mutagenesis

Several H2AFX mutants were made by site-directed mutagenesis using the pcDNA5-FRT-H2AFX_wt plasmid described in Subsection 2.5.5 as a initial template. Single nucleotide point mutations were made with 100% overlapping primers, whereas for larger deletions only partially overlapping primers were designed, as described in [318]. The PCR was performed with Q5 polymerase (New England Biolabs, #M0491L). 25 µL PCR reactions were prepared, each containing 5 µL 5x Q5 Reaction Buffer, 5 µL Q5 5x Enhancer, 0.5 µL 10 mM dNTPs (New England Biolabs, #N0447S), 0.25 µL Q5 polymerase, 100 ng template plasmid and 150 ng of each mutagenic primer. The initial denaturation was performed at 98°C for 30 sec, followed by 18 cycles of 98°C for 30 sec, annealing at the corresponding temperature for 1 min and 30 sec/kb elongation at 72°C. The final elongation step was performed at 72°C for 5 min. Table 2.5 shows the constructs with their corresponding mutagenesis primers, the template plasmids, the PCR melting temperatures and the deleted nucleotide positions. After the PCR, 10 µL of the PCR reaction were separated on a 1% agarose gel (as described in 2.12) to confirm correct PCR amplification. The remaining 15 µL of the PCR reaction were transferred to 37°C and 1 µL DpnI (New England Biolabs, #R0176S) was added to each reaction to degrade the original plasmid DNA, keeping intact the amplified DNA. After 1 hr incubation at 37°C, the tubes were incubated on ice and 5 µL were transformed in XL1-Blue *E. coli* competent cells, as described in Subsection 2.5.2.

ΔHDE clones were selected by colony PCR with *Seq Primer Fw-4* and *Seq Primer Rev-1* primers from Table 2.4, followed by BlpI restriction digestion of the PCR product. The deleted region contained a second BlpI site, therefore the PCR product was cut in 3

fragments if no deletion occured and only in two fragments if the deletion was successful. Similarly, ΔCons positive clones were selected by colony PCR with primers *Seq Primer Fw-6* and *Seq Primer Rev-1* primers from Table 2.4, and the PCR product was cut with Bsu36I. The positive clones showed only one fragment while negative clones two fragments. All positive clones found by colony PCR were sent for Sanger Sequencing (see Subsection 2.8.1). The other constructs could not be selected by colony PCR so that selection was by direct Sanger Sequencing.

#### Stable integration into HeLa-FlpIn T-REx cells

For stable integration of the construct into the HeLa-FlpIn cell line 2 µg of the FRT site-containing plasmid (see Table 2.5) together with 2 µg pOGG4 Flp-recombinase Expression Vector (Thermo Fisher Scientific, #V600520) were resuspended in 200 µL Opti-MEM (Gibco, #31985-062). 12 µL X-tremeGENE 9 DNA Transfection Reagent (Sigma, #6365787001) were added and the mix incubated at room temperature for 15-30 min. During this incubation period, $3.5 \times 10^6$ HeLa-FlpIn cells were seeded in a 10 cm dish, in 10 mL antibiotics-free media. The transfection mix was applied dropwise onto the cells and the dish was gently rotated to evenly distribute the transfection complexes.

24 hr after transfection, the media was changed to DMEM containing antibiotics and 400 µg/mL Hygromycin B (Invitrogen, #10687010) to eliminate the cells negative for the integration. Positive cells also contain the Hygromycin B resistance gene and therefore survive the drug treatment. The day after, the media was changed again and only 150 µg/mL Hygromycin B were added. After about two days, the media was changed again to eliminate dead cells and the remaining cells were grown until the appearence of colonies. Cells were then split, further expanded and frozen in multiple vials.

#### PCR and restriction digestion approach to distinguish endogenous and transgene H2AX RNA

All the integrated transgenes (see Table 2.5) were designed to contain a HindIII restriction site in the H2AFX ORF (created by inserting a silent point mutation) that is absent in the endogenous H2AFX gene. Therefore, to distinguish the RNA made from either endogenous

**Table 2.5** Primers employed for site-directed mutagenesis to create the H2AFX 3-UTR mutants for stable insertion into HeLa-FlpIn T-REx cells.

| | Construct | mutagenesis primers (name) | mutagenesis primers (sequence) | plasmid employed as template | Tm (¡C) | Mutation |
|---|---|---|---|---|---|---|
| 1 | pcDNA5-FRT-H2AFX_HindIII | H2AX-HindIII_Quik_Fw_EG<br>H2AX-HindIII_Quik_Rev_EG | GTCACGCCGCCCAGAAGCTTGTTGAGCTC<br>GAGCTCAACAAGCTTCTGGGCGGCGTGAC | pcDNA5-FRT-H2AFX_wt | 55 | nt 364 G to T |
| 2 | pcDNA5-FRT-H2AFX_HindIII_dHDE | H2AFX-dHDE-Fw-2_EG<br>H2AFX-dHDE-Rev-2_EG | CCCTCATGGAGCCGCTTCAGACTGCGGGGCAA<br>GCAGTCTGAAGCGGCTCCATGAGGGCGGTGGT | pcDNA5-FRT-H2AFX_HindIII | 65 | 555-570 |
| 3 | pcDNA5-FRT-H2AFX_HindIII_dSL | H2AFX-dSL-Fw-2_EG<br>H2AFX-dSL-Rev-2_EG | ACAAAACCACCGCCCTCATGGAAAGAGCTGAGCCGCT<br>GGGCGGTGGTTTTGTGGTGGCATGGGGAGGCCT | pcDNA5-FRT-H2AFX_HindIII | 55 | 588-596 |
| 4 | pcDNA5-FRT-H2AFX_HindIII_dConserv | H2AFX-dCons-Fw-2_EG<br>H2AFX-dCons-Rev-2_EG | TAACCCTTGGACTAATTTTATTAAAGGATTG<br>AGTCCAAGGGTTAGCTGCAGAATTCCAGTTCA | pcDNA5-FRT-H2AFX_HindIII | 55 | 1522-1558 |
| 5 | pcDNA5-FRT-H2AFX_HindIII_dSL_dHDE | H2AFX-dSL-Fw-2_EG<br>H2AFX-dSL-Rev-2_EG | ACAAAACCACCGCCCTCATGGAAAGAGCTGAGCCGCT<br>GGGCGGTGGTTTTGTGGTGGCATGGGGAGGCCT | pcDNA5-FRT-H2AFX_HindIII_dHDE | 55 | 555-570 and 588-596 |
| 6 | pcDNA5-FRT-H2AFX_HindIII_dHDE_dConserv | H2AFX-dHDE-Fw-2_EG<br>H2AFX-dHDE-Rev-2_EG | CCCTCATGGAGCCGCTTCAGACTGCGGGGCAA<br>GCAGTCTGAAGCGGCTCCATGAGGGCGGTGGT | pcDNA5-FRT-H2AFX_HindIII_dConserv | 65 | 554-570 and 1522-1558 |
| 7 | pcDNA5-FRT-H2AFX_HindIII_dsi | H2AFX-del-siRNA3-Fw_EG<br>H2AFX-del-siRNA3-Rev_EG | TCAACTCGGCAAGTCGGTTAATCCCTGTCTGGACTGAGCCTCCGTTGGCT<br>CCGACTTGCCGAGTTGAGTTTGCTGGAAGGGAAATGGGGCGGCGT | pcDNA5-FRT-H2AFX_HindIII | 72 | 1430-1454 |
| 8 | pcDNA5-FRT-H2AFX_HindIII_dsi_dPAS | H2AFX-dPAS-Fw_EG<br>H2AFX-dPAS-Rev_EG | ATGGACTAAAAATGGAGCGTTTGTTTTCATTTTGCAAGAGAGTTGTAGAATTAGTGAC<br>GCTCCATTTTTAGTCCATCTAAAACTCCCCAATGCCTAAGGTTCTAG | pcDNA5-FRT-H2AFX_HindIII_dsi | 55 | 1430-1454 and 1566-1592 |
| 9 | pcDNA5-FRT-H2AFX_HindIII_dSL_dCons | H2AFX-dSL-Fw-2_EG<br>H2AFX-dSL-Rev-2_EG | ACAAAACCACCGCCCTCATGGAAAGAGCTGAGCCGCT<br>GGGCGGTGGTTTTGTGGTGGCATGGGGAGGCCT | pcDNA5-FRT-H2AFX_HindIII_dConserv | 55 | 588-596 and 1522-1558 |

or transgene H2AFX, the RNA was transcribed to cDNA (see Section 2.9), amplified by PCR using phusion polymerase (see Section 2.10) and 50% of the PCR reaction was subjected to restriction digestion with HindIII (New England Biolabs, #R0104S). For the digest, 2 µL HindIII were added to 25 µL PCR reaction and incubated at 37°C for 1 hr. The uncut and digested PCR products were loaded on a 2% agarose gel and separated by gel electrophoresis (see Section 2.12). Table 2.6 shows the primer sequences employed to amplify either both H2AX stem-loop and polyadenylated RNA (Total) or only the polyadenylated RNA (poly(A)$^+$) and Figure 6.2 in Chapter 6 shows the resulting PCR products before and after HindIII digestion.

**Table 2.6** Primers employed to PCR-amplify transgene and endogenous H2A.X RNA from HeLa-FlpIn cell lines.

| **Primer name** | **Sequence (5'-3')** |
|---|---|
| H2A.X Total Fw-1 | CGGGCGTCTGTTCTAGTGTT |
| H2A.X Total Rev-2 | CGGGCCCTCTTAGTACTCCT |
| H2A.X polyA Fw-C | AGTGCTGGAGTACCTCACCG |
| H2A.X polyA Rev-C | GAAGGCCCGAACCCTACC |

## 2.6 CRISPR-Cas9 genome engineering

The CRISPR-Cas9 guide RNA (gRNA) as well as the template for homologous recombination were designed by Joey Riepsaame (Genome Engineering Facility, Sir William Dunn School of Pathology, University of Oxford). Target specificity of the gRNA was analysed using the CCTop online Tool (https://crispr.cos.uni-heidelberg.de). Two gRNAs were designed to cut close to the H2AFX HDE (see Table 2.7). The symmetric single-stranded donor oligonucleotide (ssODN) employed as template for homologous recombination was designed in the antisense direction of the gene, with a 26 bp left homology arm and a 29 bp right homology arm and deleting nucleotides 574-598 of the H2AFX gene (counting from the first nucleotide of the 5' UTR). An EcoRI restriction site was inserted between the homology arms, to facilitate the screening for positive clones (see Table 2.7).

The gRNAs were *in vitro*-transcribed as described in [319] and assembled with recombinant Cas9 protein. For RNP formation, 1 µg of each gRNA was added to 5 µg (HiFi) Cas9 protein (IDT, #1081060) in a total volume of 2 µL IDT Duplex Buffer (IDT,

#11-05-01-12), mixed by pipetting up and down for 5 times and incubated at 37°C for 5 min. The resulting RNPs were transfected into the cells together with 100 pmol of the template for homologous recombination by electroporation using the NEON Transfection System (Thermo Fisher Scientific, #MPK5000), as described below.

The cells were seeded in antibiotic-free media 24 hr before transfection. Two hours before harvesting for transfection, the cells were treated with the non-homologous end joining (NHEJ) inhibitor SCR-7 to push the cells towards repair by homologous recombination. For each transfection, $2 \times 10^5$ cells were employed. The cells were washed twice with PBS and nucleofected with the pre-assembled RNPs and the ssODN (described above) in a total volume of 10 µL "buffer R" (Invitrogen, #MPK1096). For HeLa cells 1005 V, 35 ms and 2 pulses were used and for HCT116 1530 V, 20 ms and 1 pulse. After transfection, the cells were immediately transferred to a 24-well plate containing 0.5 mL antibiotics-free media supplemented with 1 µM SCR-7 (Sigma, #SML1546 ). The day after the transfection, the media was changed to media with antibiotics, without SCR-7. After 48 hrs, the cells were split 1/10 into a new 24-well plate and the remaining cells were employed to seed 1 cell per well in a 96-well plate by FACS (Fluorescence-activated cell sorting) using the BD FACSAria III, each well containing 50 µL complete media. For cell sorting, DAPI was added to avoid dead cells (for the gating strategy see Supplementary Figure S1). The cells remaining after FACS sorting were employed for genomic DNA extraction (see Subsection 2.7.2) followed by bulk PCR to assess the proportion of deletion mutants versus wild-type cells.

PCR screening was performed with 5'-CGGGCGTCTGTTCTAGTGTT-3' as forward primer and 5'-ACGGAGGTCCCCGAAGAG-3' as reverse primer, resulting in a PCR product of 792 bp in the wild-type and 773 bp in the deletion mutant. To distiguish the two, the PCR product was cut with EcoRI, showing two fragments (557 bp and 216 bp) only in the EcoRI-containing deletion mutants.

For the PCR, 50 µL PCR reactions were prepared containing 300 ng genomic DNA, 0.5 µL Phusion polymerase (New England Biolabs, #M0530S), 10 µL 5x GC buffer (New England Biolabs, #M0530S), 1 µL 10 mM dNTPs (New England Biolabs, #N0447S), 2.5 µL of the 10 µM dilution of each primer and 5 µL 5 M Betaine (Sigma, #61962). The melting temperature was 66°C and the elongation time 30 sec. The remaining parameters

were set according to the manufacturer's instructions. 10 µL of each PCR reaction were kept as uncut PCR product, 20 µL were employed for digestion with EcoRI (New England Biolabs, #R0101S) and 10 µL were employed for the T7 assay described below. The PCR products were analysed on a 2% agarose gel (see Section 2.12).

Complementing the bulk PCR, a T7 assay was performed to verify the presence of mutant clones, as the bulk PCR not always showed clear smaller bands. 10 µL of the PCR reaction described above were denatured at 95°C for 5 min and then cooled to 15°C, decreasing the temperature 0.2°C/sec, to allow annealing of complementary DNA strands. 0.25 µL T7 endonuclease I (New England Biolabs, #M0302S), 2 µL NEB Buffer 2 (New England Biolabs, #B7002S) and 5.5 µL nuclease-free water (Ambion, #AM9937) were added, mixed by pipetting and incubated at 37°C for 20 min. 10 µL were separated on a 2% agarose gel (see Section 2.12).

Once the presence of potential deletion clones was determined in the bulk of cells, the corresponding 96-well plate was screened for positive clones. Each 96-well plate was split into two new 96-well plates, one with a 1/20 dilution to keep the clones growing and one with a 1/2 dilution for genomic DNA extraction. The following day, when the cells in the 96-well plate kept for genomic DNA extraction were attached to the surface of the wells, the media was poured off, cells were washed with PBS, followed by the addition of 50 µL/well lysis buffer (50 mM Tris-HCl (Sigma, #T6066), 0.5% Triton X-100) supplemented with 1 mg/mL fresh proteinase K (Sigma, #3115828001). The plates were sealed with 96-well plate sealing tape (Thermo Fisher Scientific, #15036), placed in a box with wet paper towels creating a "wet chamber" to avoid evaporation of the buffer in case any well opened, and incubated over night at 55°C. The proteinase K was heat-inactivated at 85°C for 45 sec, the plate was then cooled to room temperature and spun to eliminate any drops on the sealing tape. To ease the screening process, 5 µL of each well from the same row were pooled and 3 µL were directly employed for PCR, using the same conditions as described above. For the rows where positive clones were identified, a PCR for each of the clones was performed. In this case 10 µL of each PCR reaction were kept as uncut PCR product, 20 µL were employed for digestion with EcoRI (New England Biolabs, #R0101S) and the remaining 10 µL were kept for Sanger Sequencing (see below).

Prior to Sanger Sequencing, the PCR products were cleaned with ExoSAP-IT (Life

Technologies, #78201.1.ML). 2 µL ExoSAP-IT were added to 5 µL of the PCR reaction, mixed by pipetting and incubated at 37°C for 20 min. The reaction was stopped by heat-inactivation at 80°C for 15 min. 2.5 µL were mixed with 2.5 µL nuclease-free water (Ambion, #AM9937) and directly sent for Sanger Sequencing (see Subsection 2.8.1).

**Table 2.7 CRISPR-Cas9 guide RNA and HR template sequences.** The capital letters in the gRNA sequences correspond to the protospacer sequence. The * in the ssODN sequence (HR template) are nucleotides with a phosphorothioate bond modification. The bold letters highlight the EcoRI restriction site in the ssODN.

| Name | Sequence (5'-3') |
|---|---|
| **gRNA 1** | taatacgactcactatagGGCGGTGGTGGCCCTTAAAAgttttagagctagaaatagcaa |
| **gRNA 2** | taatacgactcactataggCTGAGCCGCTTCAGACTGCGgttttagagctagaaatagca |
| **ssODN** | G*C*GGCCCGCTTGCCCCGCAGTCTGAAGC**GAATTC**TGGTGGCCCTTAAAAGGGCCTTTG*T*G |

## 2.7 DNA purification

### 2.7.1 DNA precipitation

To precipitate DNA in solution (*e.g.* in a PCR reaction), 2.5 volumes 100% ethanol (Sigma, #32221-M) and 10% 3M sodium acetate (VWR, #27652.260) pH5.2 were added. The tube was vortexed for 10 sec and then spun at 17000 x *g* for 30 min. The supernatant was discarded and the DNA pellet was washed with 70% ethanol. After spinning at 17000 x *g* for 5 min, the supernatant was thoroughly discarded and the pellet was air-dried for 3 min. The DNA was resuspended in nuclease-free water (Ambion, #AM9937) and kept at 4°C in case of genomic DNA or -20°C for PCR products or shorter DNA fragments.

### 2.7.2 Genomic DNA extraction

For genomic DNA extraction and precipitation, $1 \times 10^6$ cells were resuspended in 400 µL lysis buffer (50 mM Tris-HCl, 0.5% Triton X-100) supplemented with 1 mg/mL fresh proteinase K (Sigma, #3115828001) and incubated at 65°C for 3 hr or 55°C over night. The proteinase K was heat-inactivated at 95°C for 5-10 min. The DNA was precipitated by adding 1 volume isopropanol (Thermo Fisher Scientific, #10588630) and spinning at 17000 x *g* for 30 min. The pellet was washed with 70% ethanol and resuspended as described above.

### 2.7.3 Plasmid DNA isolation

Plasmid DNA was isolated using the QIAprep Spin Miniprep Kit (Qiagen, #27106) for 5 mL bacterial cultures or with the Plasmid PlusMidi Kit (Qiagen, #12945) for 50-100 mL bacterial cultures, as described by the manufacturer's instructions.

## 2.8 Sequencing

### 2.8.1 Sanger Sequencing

Sanger Sequencing was performed by Source Bioscience with the samples prepared as requested by the company. The sequences were aligned to the reference sequence using the SnapGene software (see Subsection 2.5.1).

### 2.8.2 Illumina sequencing

Prior to library preparation, ribosomal RNA (rRNA) was depleted using the Ribo-Zero Magnetic Kit (Human/Mouse/Rat) from Illumina (#MRZH116). 500 ng ribo-depleted RNA were employed for library preparation with the NEBNext Ultra Directional RNA Library Prep Kit (Illumina, #E7420S). The libraries were analysed on a bioanalyser (Agilent 2100 Bioanalyzer system) using the Agilent High Sensitivity DNA Kit (Agilent, #5067-4626). 50 bp paired-end sequencing was performed by the Wellcome Trust Centre for Human Genetics (University of Oxford, UK).

#### RNA-sequencing data processing

Adapters from the RNA-sequencing reads were trimmed using Cutadapt 1.8.3. Only correctly paired and properly mapped reads were selected using SAMtools 0.1.19. The data were visualised using Bedtools 2.23.0 and scaled to the size of each library (genomeCoverageBed). The UCSC genome browser was employed to visualise the data.

## 2.9 Reverse transcription

Complementary DNA (cDNA) was made from 1-5 µg DNase-treated RNA, using random hexamers (Life Technologies, #48190011) and SuperScript IV Reverse Transcriptase (Life Technologies, #18090050), according to the manufacturer's instructions. Importantly, to transcribe the polyadenylated H2A.X RNA form, only SuperScript IV reverse transcriptase was able to transcribe over the stem loop structure (see Figure 2.6).

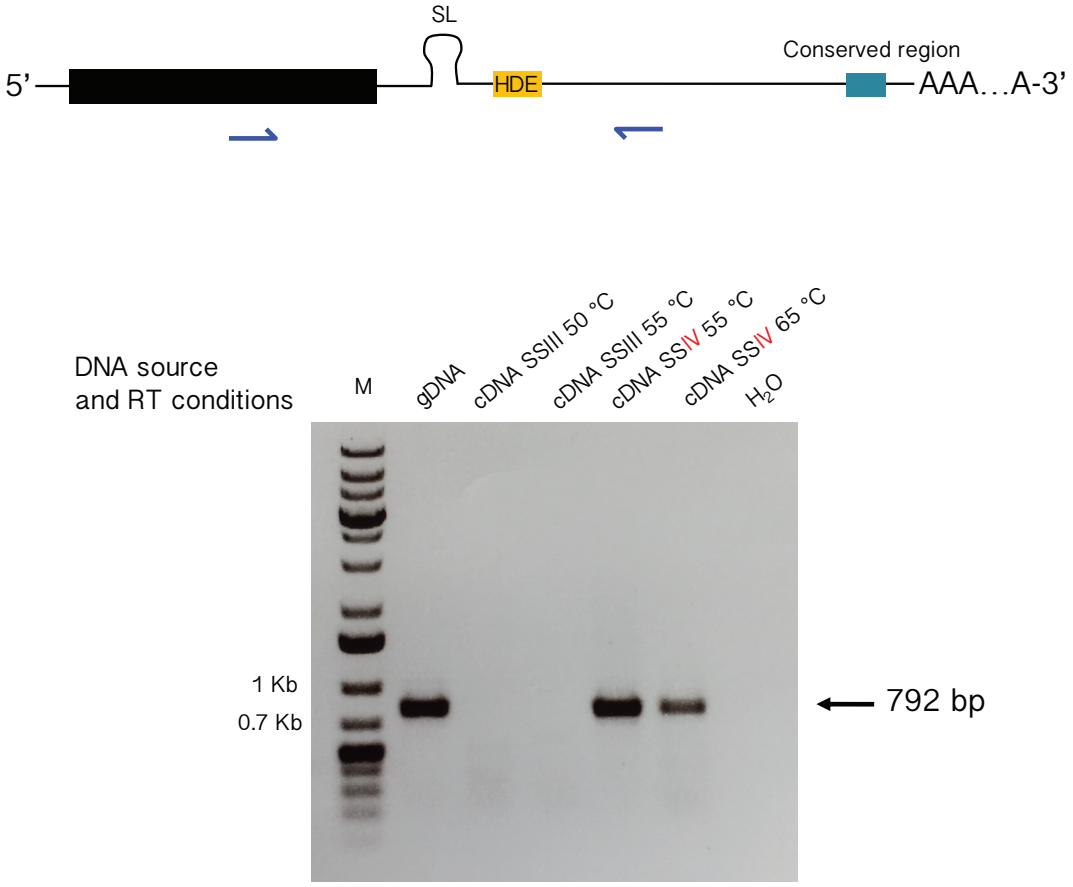



**Figure 2.6 PCR of RNA reverse-transcribed using different reverse transcriptases and different reaction temperatures.** The reverse transcription was performed using random hexamers and either SuperScriptIII or SuperScript IV reverse transcriptase, comparing different reaction temperatures. The image shows PCR products resulting from amplifying with the primers shown as blue half arrows in the H2A.X RNA schematic on the top. Genomic DNA was employed as a positive control for the PCR. Note that only reverse transcriptase SuperScript IV was able to reverse transcribe over the H2AFX SL region.

## 2.10 PCR

PCR was performed using the ProFlex PCR System (Thermo Fisher Scientific, #4484073). For site-directed mutagenesis the Q5 polymerase was employed (see Section 2.5.5) whereas for other applications, Phusion polymerase (New England Biolabs, #M0530S) was employed. The PCR reactions performed with Phusion polymerase were supplemented with 10% 5 M Betaine (Sigma, #61962). The melting temperature was determined using the NEB Tm calculator (`https://tmcalculator.neb.com/#!/main`) and the cycling parameters were set according to the manufacturer's instructions. In cases where the suggested annealing temperature was unsucessful a range of melting temperatures were tested.

## 2.11 Real-time PCR

Real-time PCR was performed using a Rotor-*GENE* Q real-time PCR cycler (Qiagen) with a 72-well rotor and 0.1 mL 4-strip Rotor-*GENE* Style tubes and caps (Starlab, #I1402-0400). 15 µL PCR reactions were prepared: 7.5 µL SYBR green (Qiagen, #204145), 0.75 µL of each 10 µM primer dilution, 4 µL nuclease-free water (Ambion, #AM9937) and 2 µL DNA (cDNA or genomic DNA, diluted accordingly). Initial denaturation was performed at 95°C for 10 min and the cycling conditions were 94°C for 20 sec, 60°C for 20 sec and 72°C for 20 sec. For the melting curve, the temperature was increased from 72°C to 95°C rising at 1°C per step, waiting for 90 sec of pre-melt conditioning on the first step and waiting for further 5 sec for each step thereafter. The source of the acquisition chanel was 470 nm and the detector 510 nm. A standard curve of 10, 2, 0.4, 0.08 and 0.016 ng DNA was loaded to calculate the absolute DNA amounts in each sample.

## 2.12 Agarose gel electrophoresis (DNA)

Agarose gels were prepared with low EEO (electroendosmosis) agarose (Melford laboratories, #A1201) in 1x TAE running buffer (40 mM Tris (Sigma, #T6066), 20 mM acetic acid (Sigma, #33209) and 1 mM EDTA (VWR, #20302.260)). The solution was heated up in the microwave until the agarose was completely dissolved. The solution was cooled to room temperature for about 10 min and then 0.2-0.5 µg/mL ethidium bromide (Sigma,

#E7637) were added for DNA detection. The gel was poured into a gel tray to solidify for about 20 min at room temperature.

To each DNA sample, DNA loading buffer was added to a final concentration of 1x (from a 6x stock containing 30% v/v Glycerol (Fisher Chemical, #G/0650/17), 0.25% w/v bromophenol blue (Sigma, #B0126) AND 0.25% w/v xylene cyanol (Sigma, #X-4126). As size marker GeneRuler 1 kb Plus DNA Ladder (Thermo Fisher Scientific, #SM1331) was employed.

The electrophoresis was performed in a horizontal electrophoresis apparatus (Gene Flow, #CSL-MultiSUB4) in 1x TAE, at 100 V constant until the desired separation was reached. The DNA was visualised using a BioDoc-It Imaging Systems (UVP, Model 97-0165-02).

## 2.13 RNA extraction

For total RNA extraction, the cells were harvested by scraping or by use of trypsin, washed with PBS and resuspended in 1 mL Tri-Reagent (Sigma,#T9424). The samples in Tri-Reagent were either stored at -80°C or further processed by adding 200 µL chloroform followed by vortexing for 10 sec and spinning for 15 min at maximum speed at 4°C. The acquose phase was transferred to a new nuclease-free 1.7 mL tube (about 500 µL) and one volume (1:1) isopropanol was added for RNA precipitation. 1 µL gycogen (GlycoBlue, Invitrogen, #AM9516) was added to samples expected to have a low RNA concentration. The tubes were inverted and vortexted for 5 sec and the RNA was precipitated by spinning at maximum speed for 30 min at 4°C. The supernatant was discarded and the RNA pellet was washed with 800 µL 70% ethanol. After spinning for 5 min at maximum speed at 4°C, the ethanol supernatant was discarded. To ensure complete ethanol removal, the tubes were spun again for 5 sec and the remaining ethanol was removed with a small pipette. The pellet was air-dried for 3 min and resuspended in $H_2O$. Depending on the downstream application, the RNA was treated with DNAse to remove any co-precipitated genomic DNA. The RNA pellet was resuspended in 200 µL DNAse buffer containing 2 µL DNAse (Turbo DNase, Thermo Fisher Scientific, #AM2238) and incubated at 37°C for 15-30 min. To eliminate any remaining DNAse, the RNA was extracted by a second

Tri-Reagent step, as described above. The RNA concentration was determined using a Nanodrop 2000 Spectrophotometer (Thermo Fisher Scientific).

## 2.14 Poly(A)$^+$ mRNA isolation

For the purification of polyadenylated RNA, the Dynabeads mRNA Purification Kit was employed (Life Technologies, #61006), according to the manufacturer's instructions.

## 2.15 RNA size selection by urea gel electrophoresis

2 µg DNase-treated total RNA (see Section 2.13) in a final volume of 4 µL were mixed with 6 µL urea loading dye (7M urea (VWR, #28877.292), 0.05% w/v xylenol cyanol (Sigma, #X-4126) and 0.05% w/v bromophenol blue (Sigma, #B0126)) and denatured at 95°C for 2 min followed by a 2 min incubation on ice. 3 µL RNA size marker (Century-Plus RNA Marker, Life Technologies, #AM7145) were mixed with 7 µL urea dye and also denatured as described above. The denatured samples were loaded on a 6% TBE urea polyacrylamide gel (Novex, %EC68652BOX) and the electrophoresis was performed in 1x TBE buffer (100 mM Tris base (Sigma, #T6066), 100 mM boric acid (Thermo Fisher Scientific, #B/3800/60) and 2 mM EDTA) at 80 V constant until the bromophenol blue dye ran out of the gel (the lower dye). To visualise the RNA, the gel was incubated in 1x TBE containing 0.2-0.5 µg/mL ethidium bromide for 15 min at room temperature or until the RNA was detectable under the UV lamp (Syngene). The desired size range (0.4 - 1 Kb and $\geq$ 1 Kb) was excised from the gel using a clean blade (Thermo Fisher Scientific, #11904325) and transferred to a nuclease-free 0.5 mL tube (Life Technologies, #AM12300). With a 25G needle (Terumo, #AN*2516R1) the bottom of the 0.5 mL tube containing the gel was perforated 3 times. The tube was transferred to a nuclease-free 1.7 mL tube (Axygen, Corning, #MCT-175-C) and spun at maximum speed for 1 min. 400 µL elution buffer (1.5 M sodium acetate (VWR, #27652.260) and 1.5 mM EDTA) were added to the fragmented gel in the 1.7 mL tube and incubated at room temperature, shaking at 1200 rpm for 1 hr (or 2 hr to elute larger sizes of RNA). The tube was spun at maximum speed for 1 min at room temperature and the supernatant was transferred to a nuclease-free filter column (Costar, #8160) in which a grade GF/D glass filter (Whatman,

#1823-010) had been previously placed. To precipitate the RNA in the flow-through, 1 mL 100% ethanol was added together with 1 µL GlycoBlue (Invitrogen, #AM9516) and spun at maximum speed for 15-20 min at 4°C. The pellet was washed with 70% ethanol and resuspended in nuclease-free water. The RNA concentration was determined using a Nanodrop 2000 Spectrophotometer (Thermo Fisher Scientific).

## 2.16 Northern blot

### 2.16.1 Agarose northern blot

For northern blotting, a horizontal gel electrophoresis apparatus with in-built buffer recirculation was employed (Thermo Fisher Scientific, #11972925). This is important to avoid the formation of a pH gradient that distorts the RNA separation. 1% agarose gels were prepared as described in Table 2.8, with low EEO agarose (Melford laboratories, #A1201), autoclaved deionised water, formaldehyde (Sigma, #F8775) and 10x MOPS (Thermo Fisher Scientific, #11889191).

**Table 2.8** Agarose gel preparation protocol for northern blotting.

| **Gel tray size (cm)** | **20 x 27** |
|---|---|
| **Gel volume** | **Vf 400 mL** |
| **Agarose** | 4 g |
| **H2O** | 222 mL |
| Microwave until agarose is completely dissolved | |
| **Weigh and add H2O to** | 337.5 g |
| **Formaldehyde 36.5-38%** | 22 mL |
| **10x MOPS** | 40 mL |
| Pour gel into the tray, remove bubbles, put combs and let solidify | |

5-10 µg RNA in a final volume of 10 µL were mixed with 6 µL RNA loading dye (50 µL formamide (VWR, #24311.260), 15 µL formaldehyde (Sigma, #F8775), 2 µL 0.5 M EDTA, 5 µL xylenol cyanol/bromophenol blue solution (0.005% xylene cyanol/0.01% bromophenol blue in water), 6 µL 10x MOPS (Thermo Fisher Scientific, #11889191) and 22 µL nuclease-free water (Ambion, #AM9937)). The RNA samples were denatured at 95°C for 5 min or at 65°C for 10 min, followed by a 2 min incubation on ice.

As running buffer, 1x MOPS was diluted from 10x MOPS (200 mM MOPS (Sigma,

#M1254, adjusted to pH7.0 with sodium hydroxide (VWR, #28244.262)), 20 mM sodium acetate (VWR, #27652.260) and 10 mM EDTA). The 1x MOPS running buffer was stored at 4°C and reused about 5 times. The electrophoresis was run at 110 V constant for about 4 hr or until the desired separation was reached.

Before the transfer, the gel was washed twice with hot deionised water (about 65°C), gently shaking for about 1 hr, to wash out the formaldehyde. This step is very important to avoid crystal formation on the membrane during the transfer due to residual formaldehyde in the gel. After the two washes, the gel was equilibrated in 10x SSC buffer (1.5 M NaCl and 150 mM tri-sodium citrate (Scientific Laboratory Supplies, #CHE3840), adjusted to pH7.0 with hydrochloric acid (Sigma, #H1758)) for about 5 min. The capillary transfer was set up as shown in Figure 2.7. The Whatman paper was soaked in 10x SSC buffer and, after placing each layer, air bubbles were removed by gently rolling a 10 mL pipette over the wet paper surface. The gel was placed on the first two layers of Whatman paper and any air bubbles were removed by gently stroking over the gel with the finger (wearing new gloves). The dry Hybond-NX membrane (Amersham, #RPN203T) was placed on the gel and any air-bubbles were removed with the pipette, as described above. The entire layer of paper towels was placed at once and on top of it a glass plate and a weight of about 500 g (*e.g.* a bottle containing 500 mL water). The transfer was performed for around 16 hr at room temperature.

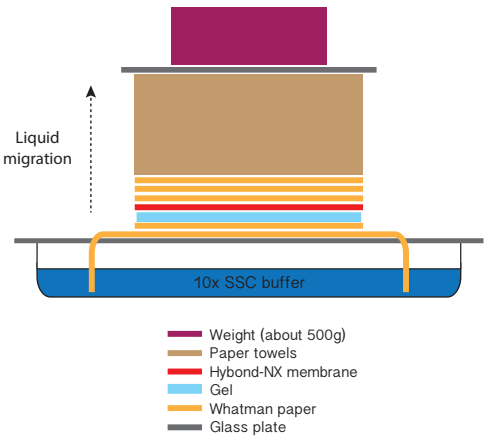



**Figure 2.7 Agarose northern blot transfer setup.**

After the transfer, the membrane was air-dried for about 10 min and then UV crosslinked with 254 nm UV light in a XL-1000 UV crosslinker (Spectronics Corporation). The upper part of the membrane, containing the RNA, was crosslinked at 1200 x100 µJ/cm$^2$ and the backside with 700 x100 µJ/cm$^2$.

After crosslinking the RNA to the membrane, the membrane was incubated with methylene blue staining solution (300 mM sodium acetate (from a 3M stock at pH5.5) and 0.1% w/v methylene blue (BDH Chemicals, #20064)) for about 5 min, gently shaking, at room temperature. Nonspecific methylene blue staining was removed by thoroughly washing the membrane with deionised water. The methylene blue-stained membrane was digitalised using a scanner (Epson Perfection V600 Photo) and later employed as loading control (specifically the 18S rRNA band).

The membrane was transferred to a hybridisation tube and 10-12 mL hybridisation buffer were added (Sigma, #H7033) followed by an incubation at 65°C, rotating for 1 hr in a hybridisation oven to allow blocking.

During the blocking incubation period, the probes were labelled. Most of the probes were dsDNA probes, generated by PCR (see Table 2.9 for primer sequences)). The PCR product was gel-purified using the Zymoclean Gel DNA Recovery Kit (Zymo Research, #D4007) and the dsDNA probes were radioactively labelled with $^{32}$P-α-dCTP (Perkinelmer, #NEG51250UC) using the Prime-IT II Random Primer Labeling Kit (Agilent, #300385), as described by the manufacturer's instructions. To detect the two H2A.X mRNA forms, a ssDNA probe was designed and ordered as custom DNA oligo with Integrated DNA Technologies (IDT). This probe was end-labelled with $^{32}$P-γ-ATP (Perkinelmer, #NEG502Z250UC) using T4 polynucleotyde kinase (New England Biolabs, #M0236S). Specifically, 50 pmol of the ssDNA oligo were resuspended in a final volume of 41 µL nuclease-free water (Ambion, #AM9937) and denatured for 5 min at 65°C, followed by an incubation on ice for 2 min. 5 µL 10x T4 PNK buffer, 2 µL $^{32}$P-γ-ATP and 2 µL T4 PNK enzyme were added (50 µL final volume), the reaction was mixed by pipetting and incubated at 37°C for 30 min. The enzyme was heat-inactivated at 65°C for 20 min. The labelled probes (dsDNA and ssDNA) were purified using Illustra Microspin G-25 columns (GE Healthcare Life Sciences, #27-5325-01), as described by the manufacturer's instructions, and denatured at 85°C for 3 min prior addition to the tube containing the

membrane and the hybridisation buffer. Probe hybridisation was performed for about 16 hr at 65°C, rotating in a hybridisation oven.

To eliminate nonspecific probe binding, the membrane was washed with warm (about 65°C) 2x SSC buffer containing 0.1% SDS. The washes were repeated until the radioactivity was undetectable in the buffer. The membrane was exposed to a Storage Phosphor Screen (GE Healthcare Life Sciences , BAS-IP MS 2025) and developed using a phosphoimager (Fujifilm, FLA-5000). Exposure times were optimised for each probe and each experiment.

### 2.16.2 Northern blot probe design

For dsDNA probes, specific PCR primers were designed using the NCBI Primer Blast online tool (`https://www.ncbi.nlm.nih.gov/tools/primer-blast/`). The specificity of the probes was first determined *in situ* by blasting the entire PCR product or the ssDNA sequence using the NCBI Blast online tool (`https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome`) and, secondly, by validating the expected RNA size observed on the northern blot. As size reference, the 18S and 28S rRNAs were employed, as well as the dyes in the RNA loading dye. See Table 2.9 for primer sequences and ssDNA probe sequences.

**Table 2.9** Primer sequences for dsDNA probes made by PCR and H2AFX ssDNA sequence. The EEF1A1 primers were designed by Silvia Ramos (University of Chapel Hill, North Carolina, USA).

| **Primers for dsDNA probes** | **Sequence (5'-3')** |
|---|---|
| ACTB-Fw | GGATTCCTATGTGGGCGACG |
| ACTB-Rev | GTAGTCAGTCAGGTCCCGGC |
| EEF1A1-Fw | TGACTGTGCTGTCCTGATTG |
| EEF1A1-Rev | CCAAAGGTGGATAGTCTGAGAAG |
| EGFP-Fw | GCAAGCTGACCCTGAAGTTC |
| EGFP-Rev | GTCTTGTAGTTGCCGTCGTC |
| HIST1H4A-Fw | TGGAGAACGTGATCCGTGAC |
| HIST1H4A-Rev | ATCAGCAACCTTAACCGCCA |
| **ssDNA northern blot probes** | |
| H2AFX | GTGGTGGCATGGGGAGGCCTGGCGGCC GGCCGCGGCGCGGGCCCTCTTAGTACT CCTGGGAGGCCTGGGTGGCCTTCTTGC CGCCCGA |

### 2.16.3 Northern blot membrane stripping

Before re-probing the membrane, the old probe was removed by incubating the membrane with boiling deionised water supplemented with 0.1% SDS, for 30 min at 65°C. To confirm the success of the stripping process, the membrane was analysed for radioactivity using the Geiger counter (Mini 900 Scintillation Monitor, Thermo Scientific). Remaining radioactivity was removed by repeating the stripping process. The clean membrane was then transferred to a hybridistation tube and incubated with 10-12 mL hybridisation buffer for blocking. The new probe was prepared as described above, in Sections 2.16.1 and 2.16.2.

### 2.16.4 Northern blot densitometry

The images obtained with the Phosphoimager were exported as .tif files and imported in the ImageStudioLite (Licor) software for densitometry analysis, as described by the manufacturer's instructions.

## 2.17 Western blot

For protein extraction with RIPA buffer, 100 µL were employed to resuspend a cell pellet of about $1 \times 10^6$ cells. 1x RIPA buffer was made from a 10x stock (0.5 M Tris-HCl pH 7.4, 1.5 M NaCl, 2.5% sodium deoxycholate (Sigma, #30970), 10% NP-40 (Sigma, #I8896), 10 mM EDTA) and supplemented with protease and phosphatase inhibitors (Thermo Fisher Scientific, #15624189) immediately before use. The cell lysate was incubated on ice for 30 min and after vortexing for 30 sec, the lysate was spun at maximum speed for 30 min. The supernatant was transferred to a new tube and either directly employed for protein quantification by Bradford (see Section 2.18) or stored at -20°C. For western blotting, 10-20 µg protein were loaded per lane.

As described above, protein extraction with RIPA buffer involves a centrifugation step that precipitates the chromatin, leading to the loss of chromatin-bound proteins, like histone proteins. Therefore, to detect total histone protein levels by western blot, the cell pellet was directly lysed in 4x SDS loading dye (400 mM Tris-HCl pH 6.8, 10% SDS w/v (Melford laboratories, #S1030), 40% Glycerol (Fisher Chemical, #G/0650/17), 4

mM EDTA, 400 mM DTT (Sigma, #10708984001) and 0.1% bromophenol blue (Sigma, #B0126)). 100 µL 4x SDS loading dye were employed to resuspend approximately $2 \times 10^6$ cells. After incubating for 15 min at 95°C to denature the proteins and fragmenting the DNA, the lysate was incubated on ice for 5 min and either stored at -80°C or diluted 1/4 with deionised water to 1x SDS loading dye. The diluted lysates were incubated at 95°C for another 5 min to ensure protein denaturation and 10 µL were loaded on the polyacrylamide gel.

As protein size marker, 5-10 µL PageRuler Plus Prestained Protein Ladder (Thermo Fisher Scientific, #26619) were loaded alongside the protein samples.

Western blots were performed using NuPAGE Novex 12% Bis-Tris Protein Gels (Life Technologies, #NP0342BOX) with NuPAGE MES SDS Running Buffer (Life Technologies, #NP0002). The electrophoresis was performed using a vertical electrophoresis system (Invitrogen, #EI0001), at a constant voltage of 200 V for about 1 hr or until reaching the desired separation.

The proteins were transferred onto a 0.2 µm nitrocellulose membrane (Amersham, #10600009) using a wet transfer apparatus (Mini Trans-Blot Cell, Bio-Rad) with 1x transfer buffer (made from 10x transfer buffer: 250 mM Tris base (Sigma, #T6066) and 1.92 M Glycin (Sigma, #G7126)) containing 20% methanol (Sigma, #32213-M). The transfer was performed at 100 V constant for 1 hr at 4°C.

The membrane containing the proteins was incubated with Ponceau S staining solution (Fluka, #09189-1L-F) for 5 min at room temperature. The membrane was washed with deionised water to eliminate nonspecific Ponceu staining and a digital image was recorded using a scanner (Epson Perfection V600 Photo).

The membrane was cut at a specific size if necessary and then incubated in TBS-T (made from 10x TBS: 500 mM Tris base (Sigma, #T6066) and 1.5 M NaCl with 0.1% Tween-20 (Sigma, #P9416)) and 3% BSA (Melford laboratories, #A1302) for 1 hr at room temperature to allow blocking. Primary antibodies were diluted in TBS-T 3% BSA and incubated over night at 4°C. The next day, the membrane was washed with TBS-T for 5 min and incubated for 1 hr at room temperature with secondary antibody in TBS-T 3% BSA (IRDye 680 RD Goat anti-mouse IgG (Licor, #926-68070) or IRDye 800 CW

Goat anti-rabbit IgG (Licor, #926-32211)). Secondary antibodies were employed at a 1:10000 dilution. See Table 2.10 for primary antibody specifications. After the incubation with the secondary antibodies, the membrane was washed twice with TBS-T for 10 min and once with TBS for 10 min. The membrane was developed using the Odyssey CLx Imaging System (Licor) and protein bands were quantified using the ImageStudioLite (Licor) software, as described in the manufacturer's instructions.

**Table 2.10** Western blot primary antibodies.

| WB Antibodies | Company | Catalog number | Description | Dilution WB |
|---|---|---|---|---|
| **Actin** | Sigma | A2066 | rabbit polyclonal | 1:10000 |
| **Fibrillarin** | Bethyl | A303-891A | rabbit polyclonal | 1:1000 |
| **Total H2A.X** | Bethyl | A300-082A | rabbit polyclonal | 1:5000 |
| **H2A.X S139 Clone JBW301** | Milipore | 05-636 | mouse monoclonal | 1:5000 |
| **H4** | abcam | ab17036 | mouse monoclonal | 1:2000 |
| **Tubulin** | Sigma | T5168 | mouse monoclonal | 1:20000 |
| **U1A** | abcam | ab166890 | rabbit monoclonal | 1:1000 |

## 2.18 Protein quantification

For protein quantification, a BSA standard curve of 5, 2.5, 1.25, 0.625 and 0.3125 µg/µL BSA (New England Biolabs, #B9001S) was prepared in RIPA buffer. 5 µL of the protein lysate or the BSA standard curve sample were resuspended in 800 µL PBS, mixed by vortexing and 200 µL Bradford reagent (Bio-Rad, #500-0006) were added. After mixing again, the reaction was incubated at room temperature for 5 min and then transferred to a cuvette (Sarstedt, #67.742) and analysed in a Spectrophotometer (WPA UV1101 Biotech Photometer).

### 2.18.1 Nuclear-cytoplasmic fractionation

80% confluent cells grown in a 150 mm dish were washed with PBS and then scraped in 5 mL PBS. After spinning at 300 x *g* for 3 min at 4°C supernatant was removed and cells were resuspended in 1 mL HLB+N buffer (10 mM Tris-HCl pH 7.5, 10 mM NaCl, 2.5 M $MgCl_2$ (AnalaR, #101494V) and 0.5% NP-40). Cells were incubated on ice for 3-10 min to allow cell swelling and plasma membrane rupture. To test if the lysis was effective, 1 µL of the lysate was placed on a slide, covered with a cover slip and analysed under the

microscope (Carl Zeiss, #491206-0007-000). After spinning at 300 x $g$ for 3 min at 4°C, the cytoplasmic lysate was transferred to a new 1.7 mL tube and either Tri-Reagent was added for RNA isolation or the lysate was directly employed for protein quantification. The nuclear pellet was resuspended once more in HLB+N buffer and carefully overlayed on 500 µL HLB+NS buffer (HLB+N buffer with 10% sucrose (VWR, #27480.294)). After pelleting the nuclei at 300 x $g$ for 10 min at 4°C, the SN was removed and the nuclear pellet was lysed either in Tri-Reagent for RNA extraction or in RIPA buffer for protein extraction.

## 2.19 Immunofluorescence

24 hr prior to the staining, $0.3 \times 10^6$ HeLa cells or $0.6 \times 10^6$ HCT116 cells were seeded on cover slips (previously sterilised by autoclaving, VWR, #631.0121). The next day, EdU was added to a final concentration of 10 µM and incubated for 30 min. The cells were carefully washed twice with PBS and immediately fixed with 4% formaldehyde (Polysciences, #18814-10) and 2% sucrose in PBS for 15 min at room temperature. After the fixation, the cells were washed 3x with PBS and then permeabilised with 0.2% Triton X-100 in PBS for 7 min at room temperature. After another wash with PBS, the cover slips were placed on parafilm in a 150 mm dish with wet paper towels placed at the border of the dish creating a wet chamber to aviod drying of the cover slips. The cells were then incubated with the EdU Click-IT reaction cocktail (Click-iT EdU Alexa Fluor 647 Imaging Kit, Thermo Fisher Scientific, #C10340) for 30 min at room temperature in the dark. All the following incubation steps were performed in the dark. Prior to the incubation with the primary antibody, the cells were washed twice with 3% BSA in PBS to block nonspecific binding. The primary antibody diluted 1:200 in complete tissue culture media (see Section 2.1.2) was incubated for 2 hr at room temperature. After 3x washes with PBS 0.01% Tween-20, the secondary antibody diluted 1:1000 in complete media was applied and incubated for 1 hr at room temperature. Following 3x more washes with PBS 0.01% Tween-20, 1:1000 DAPI staining solution was applied (Thermo Fisher Scientific, #62248). Finally, the cover slips were washed twice with miliQ water, all the excess liquid was removed using an aspiration pipette and then the cover slip was mounted on a slide (VWR, #631-0908) using ProLong Gold Antifade Reagent (Thermo Fisher Scientific, #P36930).

The cover slips were sealed with transparent nail polish and dried over night at room temperature in the dark. After that, the slides were stored at 4°C in the dark.

**Table 2.11** Primary and secondary antibodies employed for immunofluorescence.

| | Company | Catalog number | Description | Dilution WB |
|---|---|---|---|---|
| **Primary antibody** | | | | |
| H2A.X S139 Clone JBW301 | Milipore | 05-636 | mouse monoclonal | 1:200 |
| **Secondary antibody** | | | | |
| Alexa Fluor 488 | Life Technologies | Goat anti-rabbit | red | 1:1000 |

### 2.19.1 Image acquisition

Immunofluorescence imaging was performed with an Olympus FV1200 confocal laser scanning microscope. The system was equipped with a 60×/1.42-NA lens, standard photomultiplier (PMT) detectors, 655-755 nm filter (BA655-755), SDM640 and SDM490 dichroic mirrors and 405 nm, 559 nm and 635 nm lasers. The settings were adjusted to the sample with the brightest signal. All the samples from the same experiment were imaged with these same settings to be able to compare intensities between samples. Of each slide, z-stacks of 10 different fields were imaged and image processing was performed as described in Subsection 2.19.2.

### 2.19.2 Image analysis

The microscopy images were processed in the Image J software (version 2.0.0). For the projection of the images, a macro was written with close help from Lewis Arthurton (University of Oxford, UK). The first channel was DAPI (blue), the second γH2A.X (red) and the third EdU (gray). This macro takes the maximum intensity of the z-stacks and creates a .tiff file, saved in the specified directory (see Supplementary Figure S2).

Another macro was created with help from Helio Duarte Roque (Oxford University, UK) to quantify the intensity of the γH2A.X signal in the nucleus of S-phase and non-S-phase cells in an unbiased manner (see Supplementary Figure S3). Statistical analysis was performed with help from Maryam Qurashi (Oxford University, UK).

## 2.20 Pulsed-field gel electrophoresis

The cells were harvested using trypsin and washed with PBS. The cells were counted using a NC-3000 Advanced Image Cytometer (Chemometec) as described in Subsection 2.1.5. Each sample was counted twice to increase accuracy. $0.2 \times 10^6$ cells were employed to prepare agarose plugs with the CHEF Mammalian Genomic DNA Plug Kit (BioRad, #1703591), according to the manufacturer's instructions. For the pulsed-field gel electrophoresis (PFGE), a 1% agarose gel was prepared in 1x TBE (100 mM Tris base, 100 mM boric acid and 2 mM EDTA) on a levelling platform (Thermo Fisher Scientific, #12444557). Each plug was cut in half, measuring with a ruler to cut as precisely as possible and carefully positioned in the well of the agarose gel. The rest of the well was filled with 1% agarose avoiding any bubbles. The electrophoresis was performed in a pulsed-field gel electrophoresis system from Bio-Rad (CHEF MAPPER XA coupled to a Variable Speed Pump for buffer recirculation and a 1000 Mini Chiller for buffer cooling) using the settings listed in Table 2.12 and 0.5x TBE as running buffer (pre-chilled to 14°C in the electrophoresis cell). After the electrophoresis, the gel was stained in 0.5x TBE with 1µg/mL ethidium bromide for about 1 hr and the DNA was visualised with a BioDoc-It Imaging Systems (UVP, Model 97-0165-02).

**Table 2.12** PFGE electrophoresis settings programmed in the Bio-Rad CHEF MAPPER XA.

| | **Duration** | **Included angle of** | **V/cm$^2$** | **switch time** |
|---|---|---|---|---|
| Block I | 9 hr | 120° | 5.5 | 30-18s |
| Block II | 6 hr | 117° | 4.5 | 18-9s |
| Block III | 6 hr | 112° | 4 | 9-5s |

## 2.21 Polysome gradient fractionation

Polysome gradient fractionation was performed in close collaboration with the Moorman laboratory (University of North Carolina at Chapel Hill, USA) and later with the Mark Ashe laboratory (University of Manchester, UK).

### 2.21.1 Gradient preparation

The gradients were prepared 24 hr before the experiment. Gradient buffer (20 mM Tris-HCl pH7.4, 140 mM KCl (VWR, #26764.260) and 5 mM $MgCl_2$) with 10%, 20%, 30%, 40% and 50% w/v sucrose and 100 µg/mL cycloheximide (Sigma, #C7698) was prepared. With a long pipette, 2 mL of the 10% sucrose gradient buffer were transferred to an ultracentrifuge tube and the higher sucrose percentage gradient buffers were carefully underlayed, starting with the 20%. The discontinous gradients were incubated over night at 4°C to allow equilibration into a linear gradient.

### 2.21.2 Sample preparation

To ensure active proliferation on the day of the harvest, $2.5 \times 10^6$ HeLa cells were seeded in 150 mm dishes to be harvested 48 hr later at 60-70% confluency. 10 min before harvesting the cells, the media was changed to complete media containing 100 µg/mL cycloheximide. Then the cells were washed with PBS also containing 100 µg/mL cycloheximide (PBS-CHX), scraped in another 5 mL PBS-CHX and transferred to a 15 mL conical tube. The cells were pelleted at 300 x *g* and resuspended in 1 mL lysis buffer (gradient buffer with 1% Triton X-100, 10 mM DTT and 100 µg/mL cycloheximide). After incubating on ice for 10 min, the lysate was passed through a 25 gauge needle for 5 times. 100 µL were stored as *total lysate* for RNA and protein extraction. To separate the nuclei from the cytoplasmic lysate, the lysate was spun at 300 x *g* for 5 min at 4°C. The supernatant was transferred to a new 1.7 mL tube and spun at 17000 x *g* for 10 min to pellet the mitochondria. The *post-mitochondrial lysate* was transferred to a new tube and 100 µL were stored for RNA and protein extraction. The nuclear pellet was resuspended in 100 µL lysis buffer and stored for RNA and protein extraction. 700 µL of the *post-mitochondrial lysate* were loaded onto the linear 10-50% sucrose gradient (see Subsection 2.21.1). The ultracentrifuge tubes were balaned with lysis buffer, placed in a SW40 swinging bucket rotor and spun in an ultracentrifuge (Becton-Dickinson) at 32000 rpm for 2 hrs. The gradient was fractionated using a gradient fractionation system (Brandel) coupled to a continuous absorbance monitor reading at an optical density of 254 nm. RNA was extracted from each fraction using Tri-Reagent (see Section 2.13) and then employed for northern blotting

(see Section 2.19.2).

# Chapter 3

# Polyadenylation of RD histone mRNA

## 3.1 RD histone genes express polyadenylated mRNA in HeLa cells

In S-phase, 3'-end processing of RD histone pre-mRNA is very efficient so that, irrespective of the existence of a downstream cryptic polyadenylation site (PAS), most pre-mRNA processing occurs at the upstream U7-dependent site, resulting in the typical poly(A)$^-$ mRNA ending in a SL structure (hereinafter referred to as SL mRNA). In contrast, outside of S-phase, the low levels of stem-loop binding protein (SLBP) together with the exclusion of other required processing factors from the histone locus body (HLB), make U7-dependent processing very inefficient. Consequently, if a histone gene is transcribed outside of S-phase, Pol II will likely reach a cryptic polyadenylation site (PAS) with the consequent production of poly(A)$^+$ mRNA by cleavage and polyadenylation (CPA) (herinafter referred to as poly(A)$^+$ mRNA).

To detect both SL and poly(A)$^+$ mRNA populations, HeLa cells were synchronised in S-phase and G1, using aphidicolin and nocodazole, respectively (Figure 3.1A and see Chapter 2, Section 2.2, for the synchronisation protocol). Synchronisation efficiency was confirmed on a proportion of cells by propidium iodide staining followed by flow cytometry, showing 83.1% of cells in S-phase and 89.3% of cells in G1, respectively (Figure 3.1B). The arrested cells were fractionated into nuclear and cytoplasmic fractions, from which protein and RNA were isolated. Fractionation efficiency was determined by western blot

using antibodies against tubulin (exclusively cytoplasmic) and U1A (exclusively nuclear) (Figure 3.2). Notably, the nuclear RNA fraction contains nucleoplasmic RNA as well as chromatin-bound RNA. Nuclear and cytoplasmic mRNA was then subjected to poly(A)$^+$ selection, followed by rRNA-depletion of the poly(A)$^-$ mRNA. Both poly(A)$^+$ and poly(A)$^-$ mRNA were used for library preparation and sent for Illumina RNA sequencing (Figure 3.2).

The RNA was sent for Illumina sequencing with the following considerations: 1) given the different fractionation steps and the difference in volume of nucleus and cytoplasm, quantitative comparisons between samples are difficult when using the same amount of RNA for the library preparation, 2) poly(A)$^-$ RNA after rRNA depletion mainly includes poly(A)$^-$ histone mRNA, whereas poly(A)$^+$ RNA includes all poly(A)$^+$ mRNA expressed in the cell. Thus normalisation to the library size leads to substantial differences in read number between samples. However, this approach does allow qualitative comparisons, *i.e.* it allows the analysis of the presence or absence of SL or poly(A)$^+$ mRNA under each condition and which, if any, cryptic PAS is used.

RNA sequencing reads were processed as described in Chapter 2 (Subsection 2.8.2). Importantly, only uniquely aligned RNA sequencing (RNAseq) reads were used for the alignment and the reads were normalised to the library size.

All RD histone genes expressed in HeLa cells showed correctly processed SL mRNA in S-phase. Based on the presence and nucleocytoplasmic distribution of polyadenylated mRNA, histone genes were grouped into 5 categories. Category 1 corresponds to histone genes that either do not have or fail to utilise a cryptic PAS, (*i.e.* no poly(A)$^-$ mRNA detected). Category 2 histone genes have a cryptic PAS that is only inefficiently used thus producing very low amounts of poly(A)$^+$ mRNA. Thus categories 1 and 2 represent truly RD histone genes, exclusively expressing SL mRNA. Group 3 corresponds to histone genes that do express some poly(A)$^+$ mRNA but that is only detected in the nucleus. Group 4 histone genes express poly(A)$^+$ mRNA that is also exported to the cytoplasm, such as H2AFX, and category 5 includes histone genes that express spliced poly(A)$^+$ mRNA also detected in the cytoplasm. See Table 3.1 for a list of these genes and the corresponding PAS location.

Figure 3.3 shows the RNAseq reads for two examples of RD histone genes, HIST1H4A

and HIST1H3A (category 1 and 2, respectively), that only express SL mRNA and are S-phase-specific. The nuclear RNA fraction contains nucleoplasmic and chromatin-associated RNA. The reads observed in the nuclear and cytoplasmic poly(A)$^-$ G1 samples for HIST1H3A (Figure 3.3, green arrows) most likely are due to "contamination" from the 3% of S-phase cells still present in the G1-enriched cell population (Figure 3.1B). Furthermore, HIST1H3A is expressed at higher levels as compared to HIST1H4A, explaining the absence of reads in the G1 poly(A)$^-$ fraction in the latter (Figure 3.3).

HIST1H1C, HIST3H2A and HIST2H4B show examples of category 4 genes (Figure 3.4). HIST1H1C is highly expressed and thus shows the highest levels of poly(A)$^+$ mRNA compared to the other two genes. Interestingly, the HIST1H1C cryptic PAS overlaps with the HDE, resulting in a polyadenylated mRNA isoform only ∼30 nucleotides longer than the SL mRNA. HIST1H2BC is an example of a category 5 gene (Figure 3.4). It has an intron in the 3'-UTR and the spliced polyadenylated mRNA is also detected in the cytoplasm. However, the clearest example of a histone gene producing both SL and poly(A)$^+$ mRNA is H2AFX (Figure 3.5). Note that all five of these genes show poly(A)$^-$ mRNA processed at the U7-dependent site in G1 (in nuclear and cytoplasmic fractions). This is attributable to the 3% of contaminating S-phase cells in the G1-enriched cell population (Figure 3.1B), as mentioned above for HIST1H3A. Interestingly, all poly(A)$^+$ mRNA isoforms were also detected in S-phase. These transcripts may result from a combination of processing at the polyadenylation site in S-phase (this would be a small fraction because in S-phase most of the pre-mRNA is processed at the U7-dependent site), but also from the basal transcription outside of S-phase when U7-dependent processing is inefficient and all the pre-mRNA is processed at the polyadenylation site. These polyadenylated transcripts are stable throughout the cell cycle and therefore also detected in S-phase.

Taken together, more than 50% of RD histone genes in HeLa cells express a poly(A)$^+$ mRNA isoform to some extent. Importantly, this mRNA is also detected in the cytoplasm and is thus potentially translated.

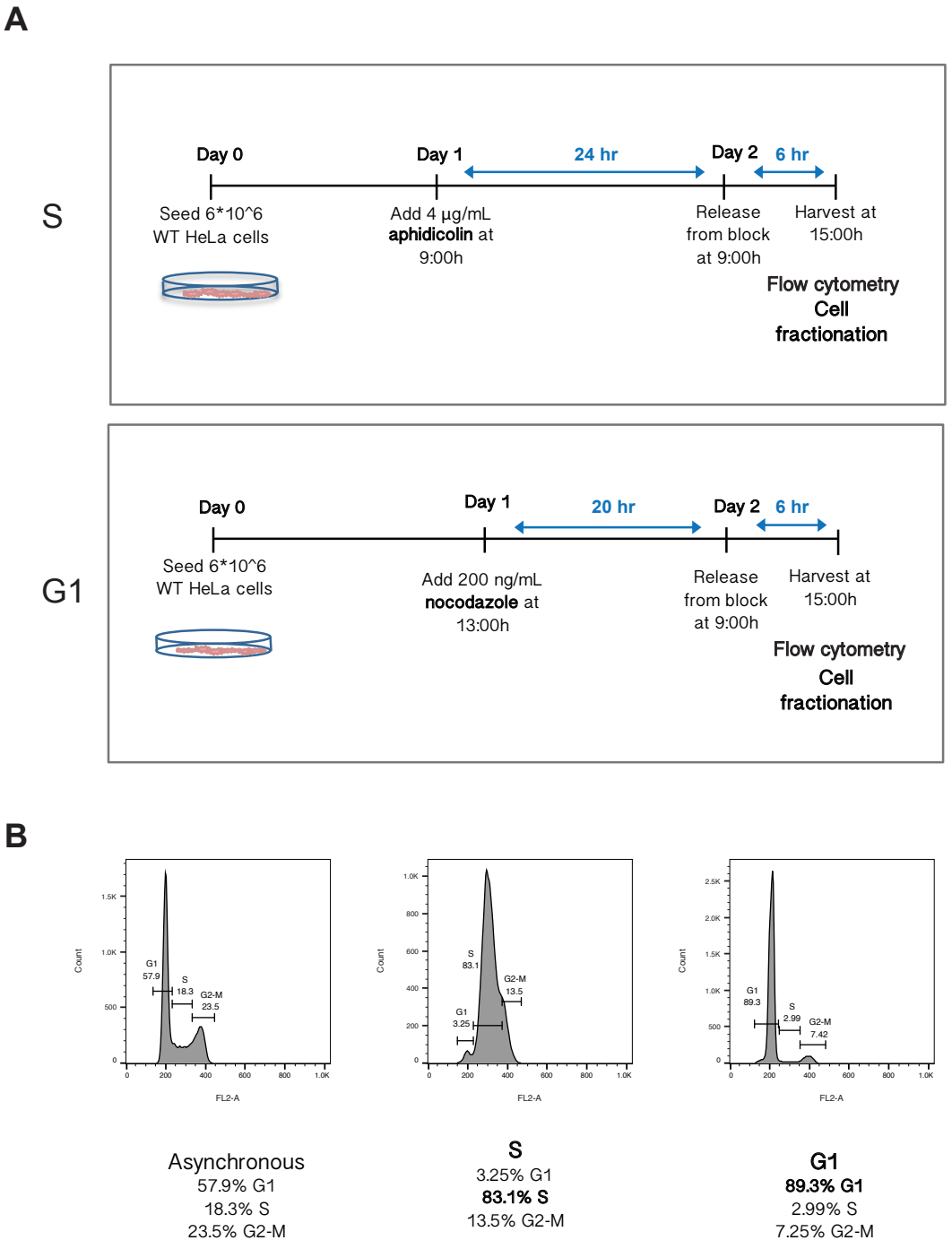



**Figure 3.1 Cell cycle synchronisation for RNAseq analysis.** A) Cell cycle synchronisation protocols. B) Flow cytometry analysis, showing the propidium iodide fluorescence on the x-axis and the cell count on the y-axis. The gating was performed manually and shows the percentage of cells in each phase of the cell cycle. The left panel shows the cell cycle distribution of asynchronously growing HeLa cells as a control, the middle panel the aphidicholin-synchronised S-phase enriched HeLa cells and the right panel the nocodazole block G1-enriched HeLa cells.

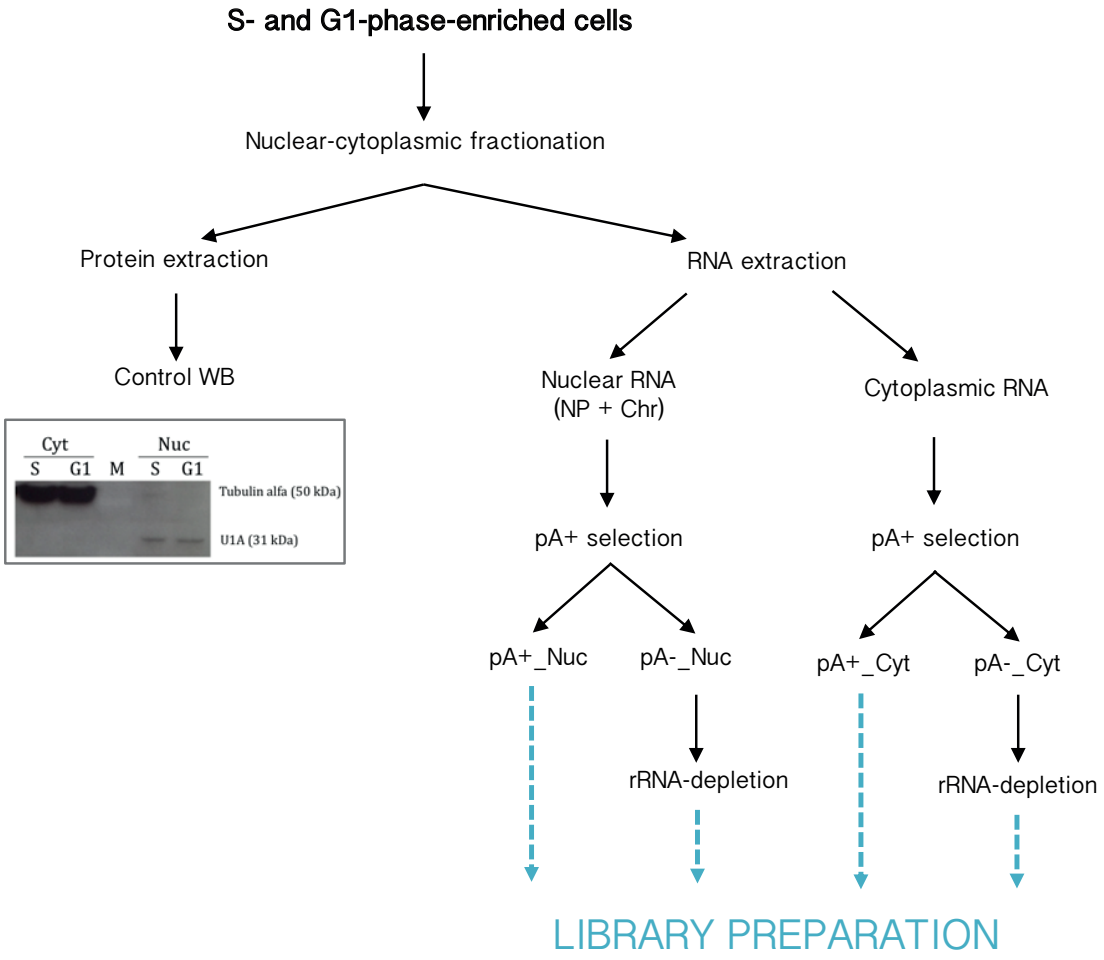



**Figure 3.2 RNA fractionation steps for RNAseq experiment.** Note that the nuclear RNA fraction contains nucleoplasmic (NP) as well as chromatin-bound (Chr) RNA.

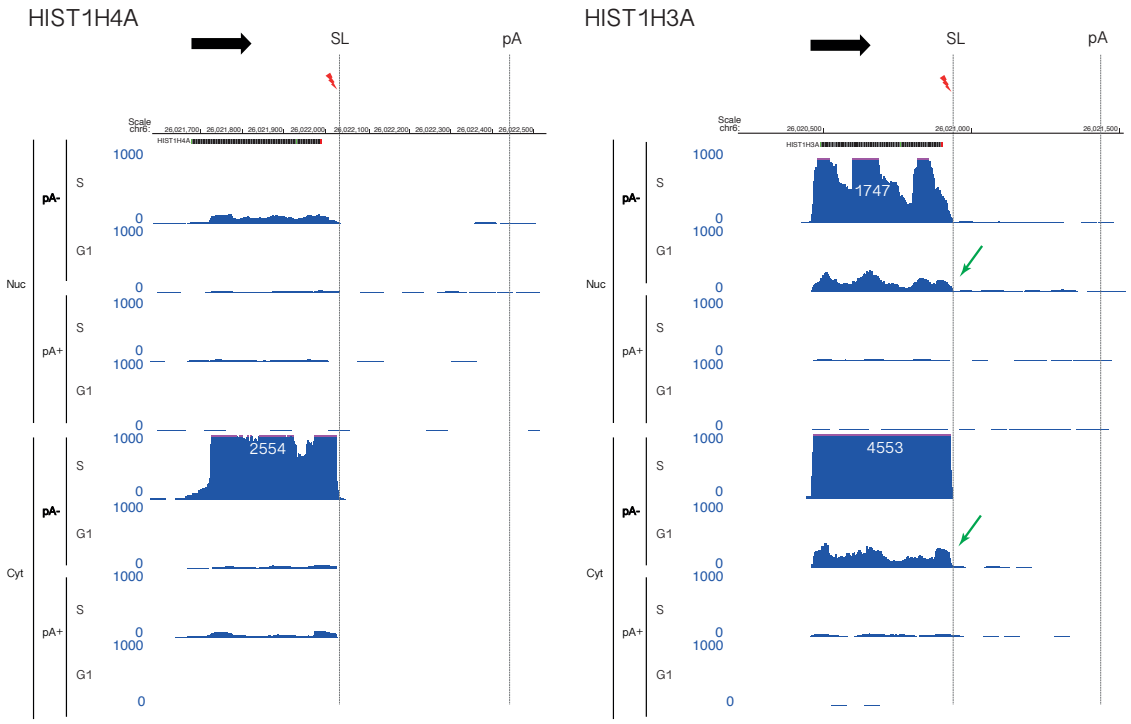



**Figure 3.3 Example RD histone genes.** RNAseq reads aligned to the human genome, visualised with the NCBI Genome Browser. The different tracks correspond to the reads from the different fractions indicated on the left. The left panel shows HIST1H4A, a typical RD histone gene from category 1 and the right panel HIST1H3A from category 2. Nuc: nuclear RNA fraction. Cyt: cytoplasmic RNA fraction. pA-: rRNA-depleted poly(A)$^-$ RNA fraction. pA+: poly(A)$^+$ RNA fraction. S: S-phase-enriched cells. G1: G1-enriched cells. SL: U7-dependent cleavage site. pA: polyadenylation site. Black arrow: direction of Pol II transcription. RNAseq reads are represented in blue because both genes are transcribed from left to right. White numbers denote maximum read number. Transcripts processed at the U7-dependent processing site observed in the poly(A)$^-$ samples likely result from the remaining 3% of S-phase cells in these samples (green arrows).

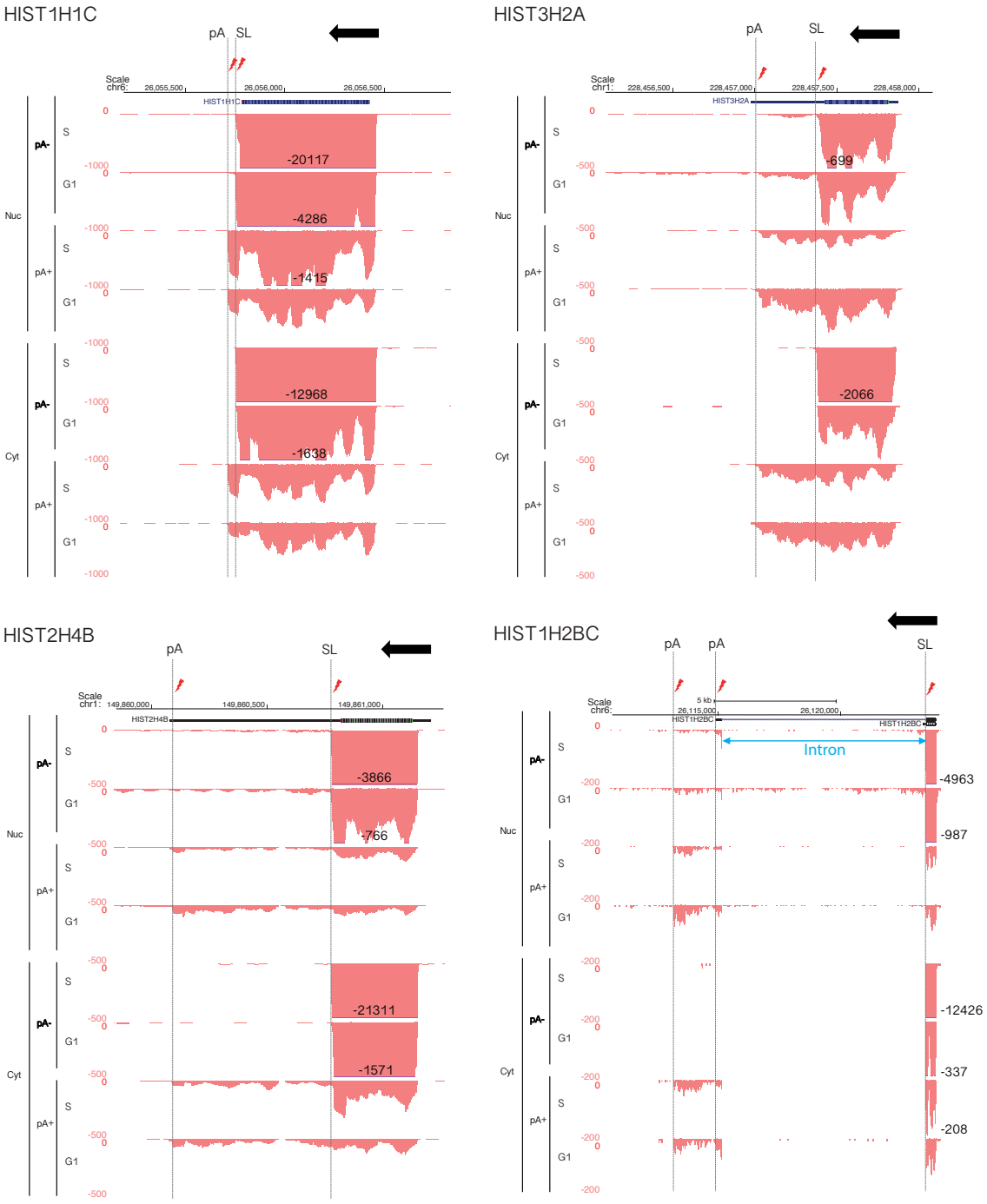



**Figure 3.4 Example RD histone genes expressing a poly(A)$^+$ mRNA isoform.** RNAseq reads aligned to the human genome, visualised with the NCBI Genome Browser. The different tracks correspond to reads from the different fractions indicated left of each panel. The two panels on the top and the bottom left panel show examples of RD histone genes from category 4: HIST1H1C, HIST3H2A and HIST2H4B. The right panel at the bottom shows HIST1H2BC, a gene from category 5. Note the intron in the 3'UTR indicated with a blue arrow. Nuc: nuclear RNA fraction. Cyt: cytoplasmic RNA fraction. pA-: rRNA-depleted poly(A)$^-$ RNA fraction. pA+: poly(A)$^+$ RNA fraction. S: S-phase-enriched cells. G1: G1-enriched cells. SL: U7-dependent cleavage site. pA: cryptic PAS. Black arrow: direction of Pol II transcription. RNAseq reads are represented in red as the genes are transcribed from right to left. Black numbers denote maximum read number.

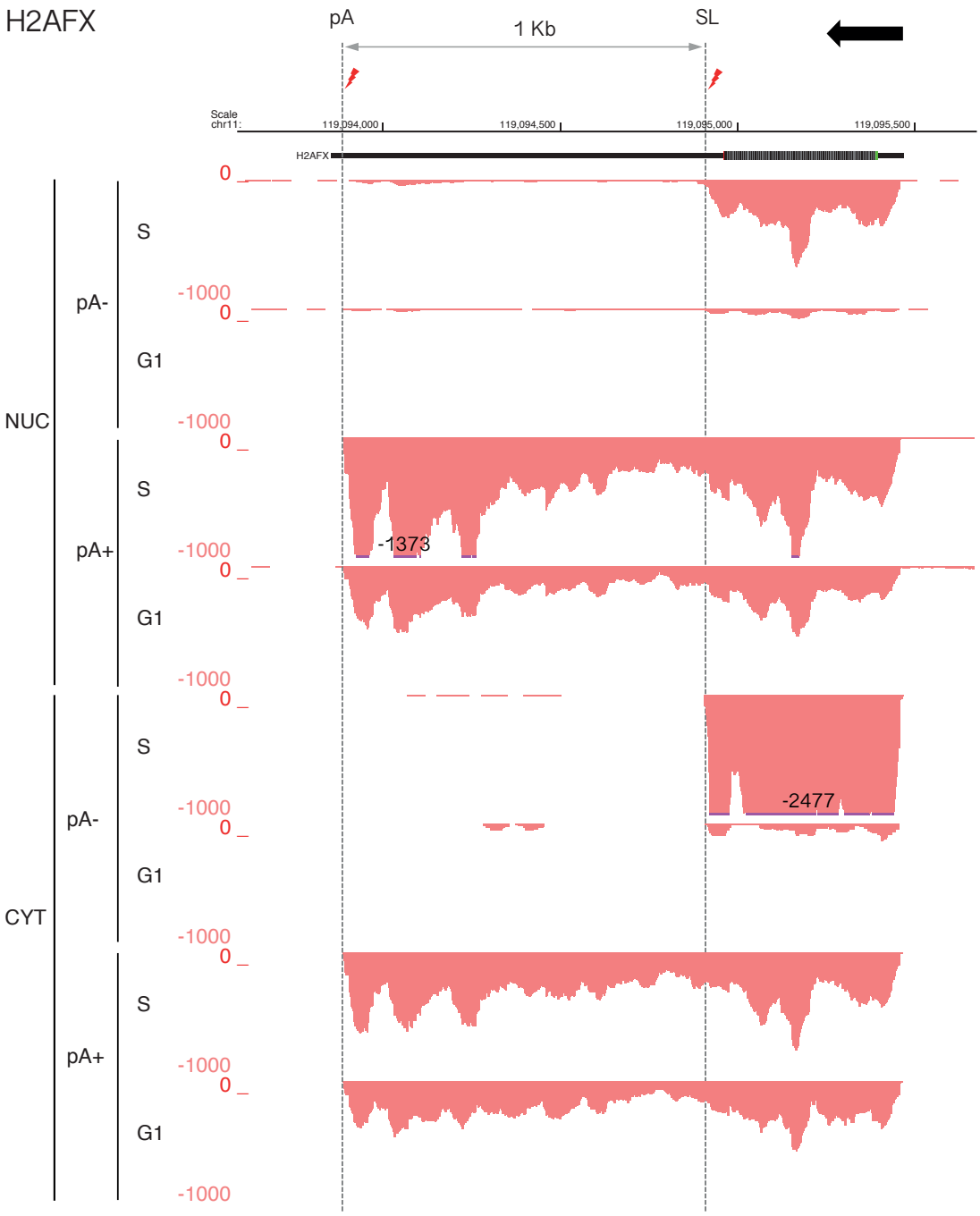



**Figure 3.5 H2AFX.** RNAseq reads aligned to the human genome, visualised with the NCBI Genome Browser. Different tracks correspond to the reads from the different fractions indicated on the left of the panel. pA-: rRNA-depleted poly(A)$^-$ RNA fraction. pA+: poly(A)$^+$ RNA fraction. S: S-phase-enriched cells. G1: G1-enriched cells. SL: U7-dependent cleavage site. pA: cryptic PAS. Black arrow: direction of Pol II transcription. RNAseq reads are represented in red because the genes are transcribed from right to left. Black numbers denote maximum read number.

**Table 3.1** Histone genes in HeLa cells.

| Gene name | Strand | Chromosome | Genomic location of PAS |
|---|---|---|---|
| **1. Histone genes only expressing SL mRNA (RD histone genes)** | | | |
| HIST1H1B | - | chr6 | unused |
| HIST1H1E | + | chr6 | in tandem gene 3'UTR ? |
| HIST1H2AD | - | chr6 | unused |
| HIST1H2AL | + | chr6 | unused |
| HIST1H2BB | - | chr6 | chr6:26041730-26041735, unused? |
| HIST1H2BI | + | chr6 | unused |
| HIST1H3C | + | chr6 | unused |
| HIST1H3J | - | chr6 | no PAS |
| HIST1H4A | + | chr6 | unused |
| HIST2H2AB | - | chr1 | unused |
| HIST2H3D | - | chr1 | unused |
| HIST3H3 | - | chr1 | unused |
| HIST1H4C | + | chr6 | unused |
| HIST1H2BE | + | chr6 | chr6:26184636-26184641, unused? |
| **2. Histone genes only expressing some poly(A)$^+$ mRNA - PAS not used efficiently** | | | |
| HIST4H4 | - | chr12 | chr12:14767659-14767664 (?) |
| HIST1H2BO | + | chr6 | unknown |
| HIST1H3A | + | chr6 | chr6:26021434-26021439 |
| HIST1H4D | - | chr6 | chr6:26186757-26186762 (?) |
| HIST1H1T | - | chr6 | chr6:26106436-26106441 |
| HIST1H2AB | - | chr6 | chr6:26032443-26032448 |
| HIST1H2AH | + | chr6 | unknown |
| HIST1H2AJ | - | chr6 | chr6:27813617-27813622 |
| HIST1H2AK | - | chr6 | chr6:27836022-27836027 |
| HIST1H2BM | + | chr6 | chr6:27817468-27817473 |
| **3. Histone genes expressing poly(A)$^+$ mRNA only detectable in the nucleus** | | | |
| HIST1H4J | + | chr6 | chr6:27825809-27825814 |
| HIST2H3A | + | chr1 | chr1:149854251-149854256 |
| HIST2H3C | - | chr1 | chr1:149839556-149839561 |
| HIST3H2BB | + | chr1 | chr1:228460308-228460313 |
| HIST1H3F | - | chr6 | chr6:26248584-26248589 |
| HIST2H2AC | + | chr1 | possibly chr1:149888153-149888158 |
| HIST2H2BF | - | chr1 | chr1:149805890-149805895 |
| HIST1H3E | + | chr6 | chr6:26227452-26227457 |
| **4. Histone genes expressing poly(A)$^+$ mRNA that is exported to the cytoplasm** | | | |
| H2AFX | - | chr11 | chr11:119093893-119093898 |
| HIST1H1C | - | chr6 | chr6:26055726-26055731 |
| HIST2H2AA3 | - | chr1 | chr1:149842204-149842209 |
| HIST2H2AA4 | + | chr1 | chr1:149851603-149851608 |
| HIST1H2AG | + | chr6 | chr6:27135269-27135274 |
| HIST1H2BF | + | chr6 | chr6:26202679-26202684 |
| HIST1H2BG | - | chr6 | chr6:26214086-26214091 |
| HIST1H2BL | - | chr6 | chr6:27805608-27805613 |
| HIST1H3H | + | chr6 | chr6:27811278-27811283 |
| HIST2H2BE | - | chr1 | chr1:149884477-149884482 |
| HIST2H4B | - | chr1 | chr1:149860100-149860105 |
| HIST3H2A | - | chr1 | chr1:228456997-228457002 |
| HIST1H2AI | + | chr6 | chr6:27809882-27809887 |
| HIST1H2BN | + | chr6 | chr6:27855557-27855562 (possibly multiple) |

| | | | |
|---|---|---|---|
| HIST1H3B | - | chr6 | chr6:26030082-26030087 |
| HIST1H3G | - | chr6 | chr6:26269421-26269426 |
| HIST2H4A | + | chr1 | chr1:149834607-149834612 |
| HIST1H2AM | - | chr6 | chr6:27891259-27891264 (?) |
| HIST1H2BH | + | chr6 | chr6:26253689-26253694 |
| HIST1H4E | + | chr6 | chr6:26206016-26206021 |
| HIST1H4K | - | chr6 | unknown |
| HIST1H4I | + | chr6 | chr6:27140615-27140620 |
| HIST1H4B | - | chr6 | unknown |

**5. Histone genes expressing spliced poly(A)$^+$ mRNA that is exported to the cytoplasm**

| | | | |
|---|---|---|---|
| HIST1H2BC | - | chr6 | chr6:26113185-26113190 |
| HIST1H2AC | + | chr6 | chr6:26138935-26138940 |
| HIST1H2BD | + | chr6 | chr6:26171325-26171330 |
| HIST1H2AE | + | chr6 | chr6:26218348-26218353 |
| HIST1H4H | - | chr6 | chr6:26277620-26277625 / 26281074-26281079 |
| HIST1H2BJ | - | chr6 | chr6:27125912-27125917 (further downstream?) |
| HIST1H3D | - | chr6 | chr6:26194219-26194224 |
| HIST1H2BK | - | chr6 | chr6:27138313-27138318 |

## 3.2 H2A.X mRNA cell cycle regulation

After identifying which RD histone genes express polyadenylated mRNA, the aim of this thesis was to determine the cell cycle regulation as well as the function of these mRNA isoforms. The RNAseq analysis identified H2AFX as a good candidate for further analysis. It clearly produces both SL and poly(A)$^+$ mRNA, but also because, unlike HIST1H1C, the H2A.X poly(A)$^+$ mRNA isoform is ∼1 Kb longer than the H2A.X SL mRNA, enabling the distinction of the two mRNA isoforms by RNA detection techniques such as northern blot. Furthermore, the H2A.X protein can be specifically detected with antibodies, which is not the case for the protein generated by other specific RD histone genes that generate poly(A)$^+$ mRNA, as there are several genes encoding for the same protein.

To analyse the cell cycle distribution of the two H2A.X mRNA isoforms, HeLa cells were synchronised using a double thymidine block (DTB), and the cell cycle profile was analysed over a period of 16 hr following release from the second block. To precisely quantify the number of cells in S, G1 or G2/M, the cells were labelled with EdU and stained with DAPI followed by flow cytometry (see Figure 3.6A for DAPI profiles and Supplementary Figure S4 for the EdU labelling). EdU (5-Ethynyl-2'-deoxyuridine) is a thymidine analogue that is incorporated into the DNA during replication thus exclusively labelling S-phase cells when applied for a short period of time (30 min). DAPI (4', 6-diamidino-2-phenylindole) unspecifically binds to the minor grove of DNA but shows a preference for AT-rich regions. Thus DAPI marks the nuclei of all the cells, independently of the cell cycle phase. Equal amounts of total RNA of each time point were analysed by northern blotting. As shown in Figure 3.6B, the H2A.X SL mRNA is strictly S-phase specific, in a similar manner to canonical RD H4 SL mRNA (referred to as H4 SL - the probe was designed for HIST1H4A but cross-hybridises to SL mRNA from other H4 genes due to the high sequence similarity). In contrast, H2A.X poly(A)$^+$ mRNA is detected throughout the cell cycle, peaking in G2/M. The mRNA encoding for the eukaryotic translation elongation factor 1 alpha 1 (EEF1A1) was used as a loading control, as was the methylene blue-stained 18S rRNA. As shown in Figure 3.6C, H4 SL and H2A.X SL mRNA levels correlate with the percentage of cells in S-phase. The H4 SL mRNA shows a stronger upregulation in S-phase compared to the H2A.X SL mRNA, matching the weaker linkage of H2AFX transcription to DNA replication, as previously suggested [255, 256].

The H2A.X mRNA isoforms showed the same cell cycle distribution in HCT116 cells (Figure 3.7). However, the S-phase duration is shorter in HCT116 cells compared to HeLa cells (Figure 3.7A), which is also reflected in the expression of H2A.X SL and H4 SL mRNA (Figure 3.7B). In this experiment, instead of measuring the percentage of S-phase cells by EdU labelling and flow cytometry, a western blot for S-phase-specific SLBP was performed (Figure 3.7C).

An interesting observation in both HeLa and HCT116 DTB-synchronised cells is that the H2A.X SL mRNA is detected at the 0 hr time point whereas the H4 SL mRNA is not yet expressed (Figures 3.6B and 3.7B). At 0 hr the cells are at the end of the second thymidine block, arrested at the G1/S-phase border. At that time SLBP (and possibly other factors required for U7-dependent processing) is already upregulated (Figure 3.7C). This observation led to the hypothesis that the H2AFX gene is already actively transcribed before S-phase is initiated. As H2AFX is located outside of the general histone gene cluster, the transcriptional attenuation still occuring in the HLB around the histone gene clusters before starting DNA replication (*i.e.* at 0 hr when the cells are at the G1/S border) may not affect H2AFX gene expression. This would result in processing at the U7-dependent processing site and thus production of H2A.X SL mRNA before DNA replication is initiated. *I.e.* the production of H2A.X SL mRNA depends on the availability of histone 3'-end processing factors, such as SLBP and active U7 snRNP, whereas additional activation mechanisms are required to activate transcription and mRNA processing of the histone genes in the cluster (*e.g.* the HLB might more stringently regulate U7 snRNP activity not affecting the U7 snRNP activity in the cytoplasm).

This hypothesis was tested on retinal pigmemt epithelium cells (RPE-1). These are human epithelial cells immortalized with human telomerase reverse transcriptase (hTERT). RPE-1 cells are thus non-transformed cells that are closer to normal cells, as is their cell cycle and the additional regulation of the transition from the resting (G0) to the proliferating state, including G1 and the G1 to S-phase transition. These cells undergo contact inhibition (arrest in G0), but re-enter the cell cycle when re-seeded at lower density. As in G0 phase of the cell cycle cells experience a general histone gene transcription shut-off, this allowed monitoring the onset of transcription starting at G1 and progressing into S-phase. RPE-1 cells were arrested in G0 by contact inhibition for four days and then

seeded at low density to allow cell cycle re-entry, from G0 to G1 and eventually to S-phase. As shown in Figure 3.8A, at 12 hr the cells were still in G1 (possibly a significant fraction still in G0) whereas at 22 hr 46% of the cells had entered S-phase. The accumulation of H2A.X poly(A)$^+$ mRNA at the 12 hr time point demonstrates that the H2AFX gene is indeed transcribed in G1 (Figure 3.8B and C). In the unaltered cell cycle, SLBP is only upregulated when the cells enter S-phase, thus in G1, the H2A.X pre-mRNA is mainly processed at the polyadenylation site by CPA, giving rise to H2A.X poly(A)$^+$ mRNA. However, the H2A.X poly(A)$^+$ mRNA continues to increase together with the H2A.X SL mRNA as the cells progress into S-phase (22 hr). There are two non-mutually exclusive explanations for this effect. One possibility is that the U7-dependent processing at the H2AFX locus is not as efficient in the S-phase as it is in the HLB at the histone gene clusters, allowing Pol II transcription past the SL to the downstream PAS and consequent processing by CPA generating H2A.X poly(A)$^+$ mRNA. The other possibility is that recovery from G0 arrest is not immediate. The ∼50% of cells that are in G1 or still recovering from G0 at the 22 hr time point continue activating H2AFX transcription so that the H2A.X poly(A)$^+$ mRNA observed at 22 hr corresponds to H2A.X poly(A)$^+$ mRNA produced in G1 cells. Possibly a combination of these two options exists, as in HeLa and especially in HCT116 cells the H2A.X poly(A)$^+$ mRNA also appears somewhat increased as the cells progress from 0 to 3 hr after release from the DTB (Figures 3.6B and 3.7B). This time frame corresponds to the activation of DNA replication, so that this increase in H2A.X poly(A)$^+$ mRNA levels most likely originates in S-phase cells.

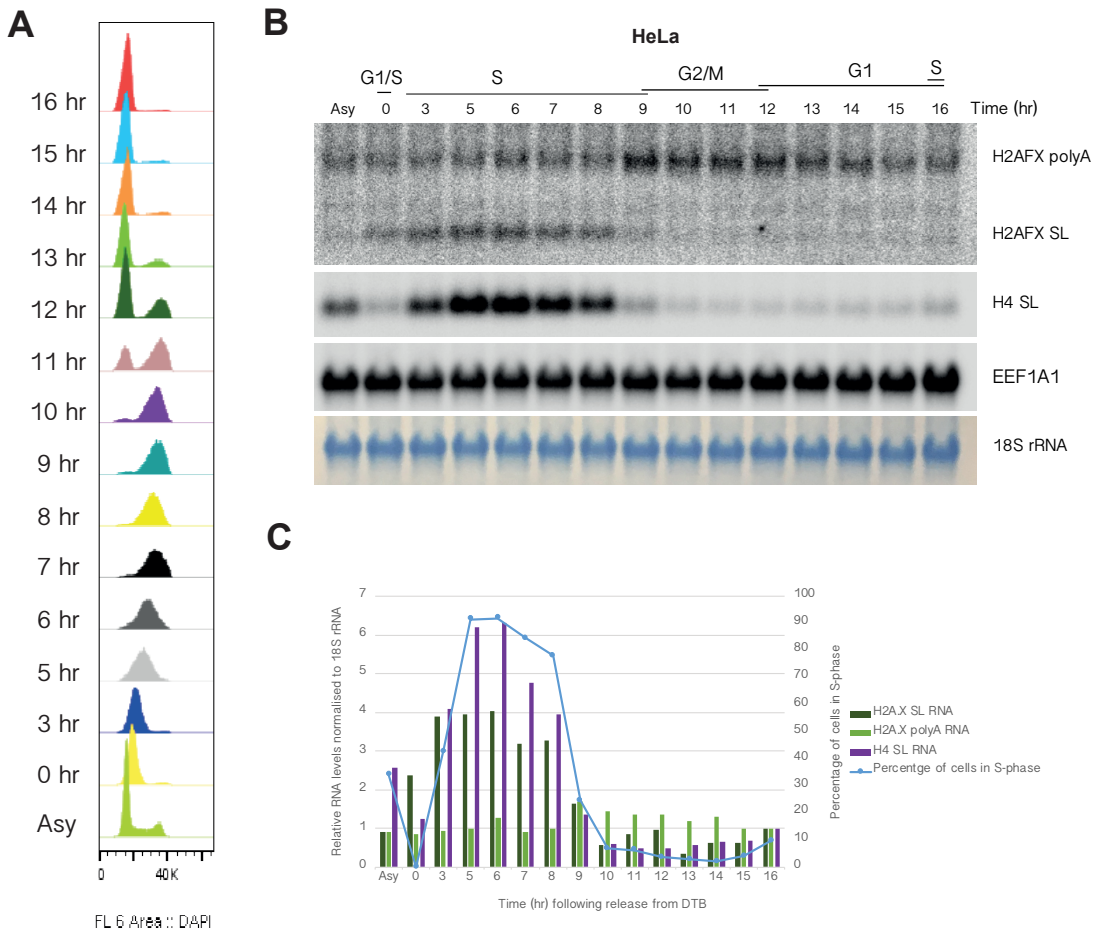



**Figure 3.6 H2A.X mRNA cell cycle distribution in HeLa cells.** HeLa cells were synchronised by DTB and time points taken after release from the second block. A) DAPI cell cycle profiles showing the DAPI fluorescent intensity on the x-axis and the cell count on the y-axis. B) Total RNA northern blot. Asy: asynchronously growing cells. The time points correspond to the time in hours following the release from DTB. The corresponding cell cycle phase determined by flow cytometry is indicated above the time points. H2A.X SL mRNA and H2A.X poly(A)$^+$ mRNA isoforms were detected with a probe complementary to the common region between both mRNA isoforms. H4 SL: canonical H4 SL mRNA. EEF1A1: eukaryotic translation elongation factor 1 alpha 1. 18S rRNA: 18S ribosomal RNA (methylene blue staining). C) Densitometry of northern blot bands shown in B compared to the percentage of S-phase cells determined by EdU labelling followed by flow cytometry. The x-axis shows the different time points in hours following release from the second block. The left y-axis shows the H2A.X SL (dark green), H2A.X poly(A)$^+$ (light green) and H4 SL (purple) RNA levels normalised to the 18S rRNA levels and represented relative to the 16h time point. The right y-axis shows percentage of S-phase cells at each time point (blue dotted line).

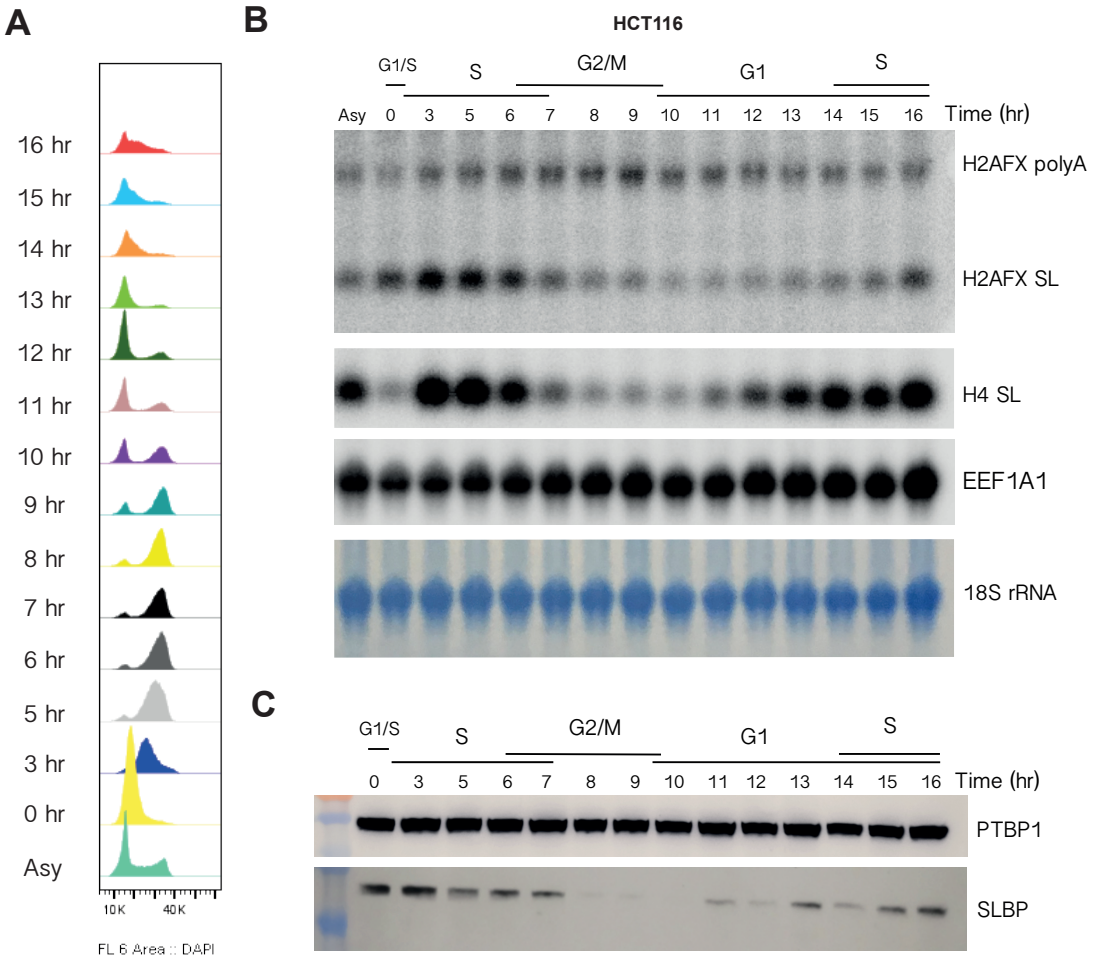



**Figure 3.7 H2A.X mRNA cell cycle distribution in HCT116 cells.** HCT116 cells were synchronised by DTB and time points were taken after the release from the second block. A) DAPI cell cycle profiles showing the DAPI fluorescent intensity on the x-axis and the cell count on the y-axis. B) Total RNA northern blot. Asy: asynchronously growing cells. The time points correspond to time in hours following release from the DTB. H2A.X SL mRNA and H2A.X poly(A)$^+$ mRNA isoforms were detected with a probe complementary to the common region between both mRNA isoforms. H4 SL: canonical H4 SL mRNA. EEF1A1: eukaryotic translation elongation factor 1 alpha 1. 18S rRNA: 18S ribosomal RNA (methylene blue staining). C) Western blot. SLBP: stem-loop binding protein. PTBP1: polypyrimidine tract-binding protein 1. The corresponding cell cycle phase determined by flow cytometry is indicated above the time points in both B and C.

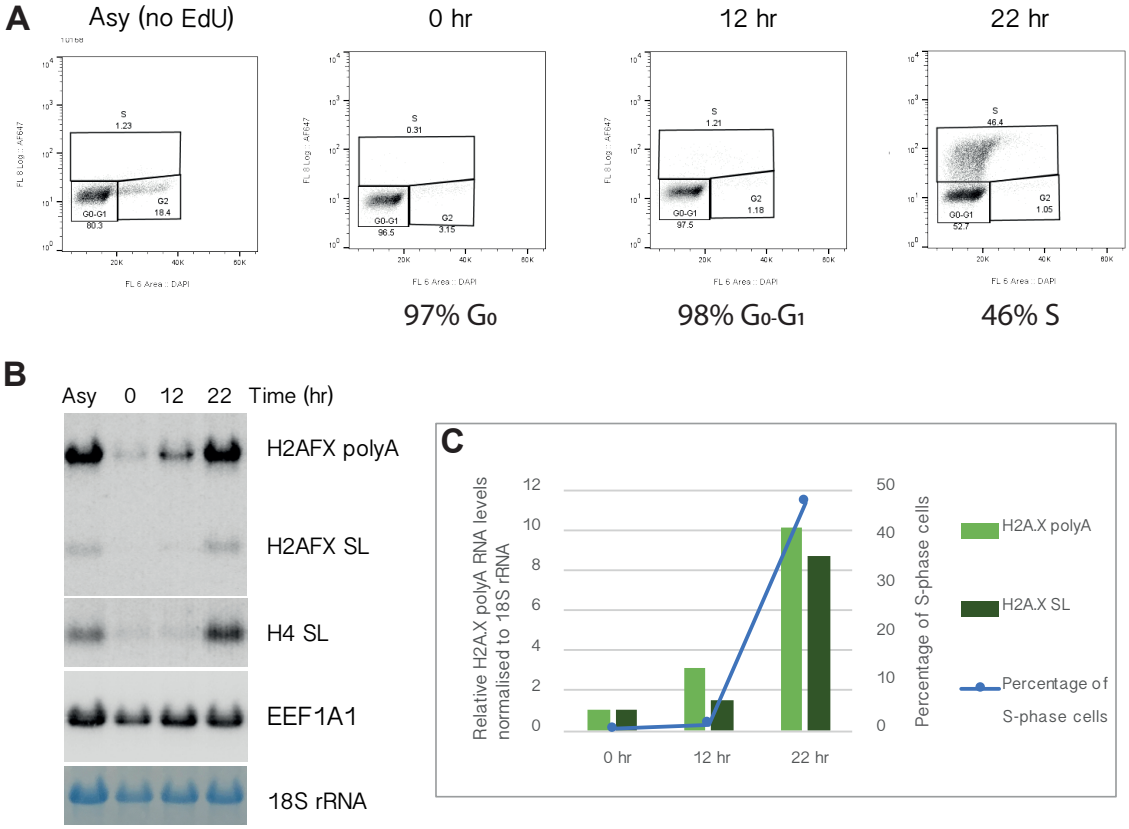



**Figure 3.8 H2A.X mRNA expression in RPE-1 cells.** RPE-1 cells were arrested by contact inhibition and harvested at different time points after cell cycle re-entry (indicated in hr). A) EdU labelling showing the DAPI fluorescent intensity on the x-axis and Alexa 647-EdU fluorescent intensity on the y-axis. Percentages of cells in G1, S and G2/M were determined using the FlowJo software. B) Total RNA northern blot (representative blot of three independent experiments). H2A.X SL mRNA and H2A.X poly(A)$^+$ mRNA isoforms were detected with a probe complementary to the common region between both mRNA isoforms. H4 SL: canonical H4 SL mRNA. EEF1A1: eukaryotic translation elongation factor 1 alpha 1. 18S rRNA: 18S ribosomal RNA (methylene blue staining). C) Northern blot densitometry. The x-axis shows the different time points after the release from the G0 arrest. The left y-axis shows the H2A.X poly(A) (light green) and H2A.X SL (dark green) RNA amounts normalised to 18S rRNA levels and relative to the 0h time point. The right y-axis represents the percentage of cells in S-phase (blue dotted line). Asy: asynchrounous cells.

# Chapter 4

# H2A.X poly(A)$^+$ mRNA

## 4.1 H2A.X poly(A)$^+$ mRNA depletion

H2A.X poly(A)$^+$ mRNA is ∼1 Kb longer than H2A.X SL mRNA. Consequently, the 3'-UTR extension can be targeted with siRNA to specifically deplete the H2A.X poly(A)$^+$ mRNA. To minimise off-target effects, four siRNAs were designed and used in combination (herinafter referred to as "siRNA pool"). However, in parallel a fifth "single siRNA" was used as a control (Figure 4.1A). The required siRNA amount as well as the time for complete depletion was tested in HeLa cells by real-time PCR (data not shown). These results showed that 10 nM siRNA are sufficient and that the mRNA depletion takes up to 24 hr. However, the protein was only substantially depleted after 48 hr (data not shown). Therefore, all the H2A.X poly(A)$^+$ mRNA depletion experiments were performed after a minimum of 48 hr. Apart from these optimisation steps, the "knock down" efficiency was always tested by northern blotting (Figure 4.1B). A control siRNA targeting the in humans non-existent luciferase transcript was also employed (siLuc). It was immediately apparent that in HeLa cells, the depletion of the H2A.X poly(A)$^+$ mRNA leads to a substantial depletion of the H2A.X protein levels (Figure 4.1C), indicating that this isoform is translated and important to maintain H2A.X protein levels. Interestingly, H2A.X poly(A)$^+$ mRNA depletion also leads to a decrease in H2A.X SL mRNA and H4 SL mRNA (Figure 4.1B). This could indicate a defect in cell cycle distribution, as has been observed before when the H2A.X poly(A)$^+$ mRNA was depleted by miRNA-138 [313]. Indeed, already after a 48 hr depletion, the percentage of S-phase cells is decreased compared to the siLuc control

(Figure 4.1D). Instead, the cells accumulate in G1.

To further test this cell cycle effect, RPE-1 cells were transfected with either siLuc or the siRNA pool and arrested in G0 by contact inhibition for four days. The cells were then diluted to allow cell cycle re-entry and harvested at different time points to monitor S-phase entry. The percentage of S-phase cells was substantially decreased in cells depleted of the H2A.X poly(A)$^+$ mRNA (Figure 4.2A and Supplementary Figure S5), suggesting a failure or delay in S-phase entry. Consistent with this observation, the northern blot of this experiment shows less H2A.X SL and H4 SL mRNA accumulation in the H2A.X poly(A)$^+$ mRNA depleted samples compared to the siLuc transfected cells as the cells progress into S-phase (Figure 4.2B).

Given the role of H2A.X in the DNA damage response, the decrease in S-phase cells observed upon depletion of the H2A.X poly(A)$^+$ mRNA may be caused by an accumulation of unrepaired DNA damage that prevents or delays S-phase entry (*i.e.* initiation of DNA replication). DNA damage induced by treatment with the topoisomerase II inhibitor etoposide [320] does not lead to a compensation of H2A.X poly(A)$^+$ mRNA levels in cells depleted of this mRNA isoform, in RPE-1 and HeLa cells (Figure 4.2B and 4.3B). Therefore, these cells may not be able to generate the required amount of H2A.X protein for efficient DNA damage repair. Pulsed-field gel electrophoresis was used to confirm that in RPE-1 cells, H2A.X poly(A)$^+$ mRNA depletion indeed leads to the accumulation of fragmented DNA, caused by unrepaired DSBs (Figure 4.2C, compare siLuc and siRNA pool in 0 and 8 hr time points marked by red arrows). Etoposide treatment for 2 hr in G1 (from 6 hr to 8 hr) increases the accumulation of fragmented DNA even further (Figure 4.2C, compare siLuc and siRNA pool in the 8 hr+E sample, marked by an orange arrow). The 22 hr time point shows the reverse effect (Figure 4.2C, compare siLuc and siRNA pool in the 22 hr samples, marked by green arrows). At 22 hr, ∼40% of the siLuc cell population corresponds to S-phase cells, whereas only ∼20% are in S-phase in the H2A.X poly(A)$^+$ mRNA depleted cell population (Figure 4.2A). S-phase cells are subjected to DNA damage accumulation due to replication stress, therefore the higher the percentage of cells in S-phase, the higher the accumulated DNA damage. As the siLuc cell population contains twice as many cells in S-phase, this may explain the higher accumulation of DNA fragments in siLuc 22 hr samples. Sorting the cells of the 22 hr time point into S-phase

and G1-phase cells by FACS followed by pulsed-field gel electrophoresis will be required in the future to detect the DNA damage accumulation in G1 cells that are depleted of the H2A.X poly(A)$^+$ mRNA. Also note that the H2A.X poly(A)$^+$ mRNA siRNA depletion efficiency drops at 22 hr, which might allow for sufficient H2A.X protein production thus explaining why the H2A.X and H4 SL mRNA levels do not vary between siLuc and siRNA pool at this time point (Figure 4.2C, compare siLuc and siRNA pool of the 22 hr time point without etoposide).

Taken together, the H2A.X poly(A)$^+$ mRNA is required to maintain H2A.X protein levels in HeLa cells and it is important for normal cell cycle progression in both HeLa and RPE-1 cells. Experiments performed in the latter suggest that the H2A.X poly(A)$^+$ mRNA may be important to maintain genome integrity.

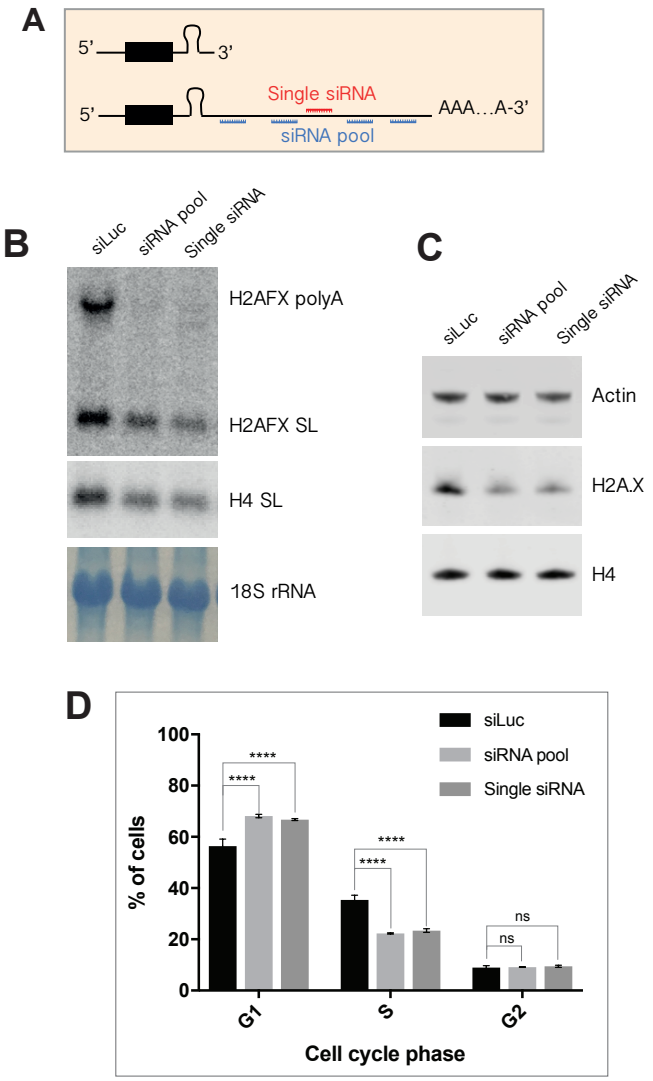



**Figure 4.1 H2A.X poly(A)$^+$ mRNA depletion in HeLa cells.** The H2A.X poly(A)$^+$ mRNA "knock down" was performed using either a pool of 4 siRNAs (siRNA pool) or a single siRNA targeting the extended 3'-UTR (single siRNA). A) Schematic representation of the siRNA binding positions. B) Total RNA northern blot of HeLa cells treated with the corresponding siRNAs for 72 hr. H2A.X SL mRNA and H2A.X poly(A)$^+$ mRNA were detected with a probe complementary to the common region between both mRNA isoforms. H4 SL: canonical H4 SL mRNA. 18S rRNA: 18S ribosomal RNA (methylene blue staining). C) Total protein extract western blot of HeLa cells treated with the corresponding siRNAs for 72 hr. Actin and the canonical histone H4 were used as controls. H2A.X indicates total H2A.X protein levels. D) Percentages of cells in G1, S and G2/M analysed by flow cytometry in HeLa cells treated with the corresponding siRNAs for 48 hr. The staining for flow cytometry was performed using propidium iodide. The error bars represent the standard error from three independent experiments, statistical significance was calculated by Two-way Anova, ns: not significant P-value $> 0.05$, **** P-value $\leq 0.0001$.

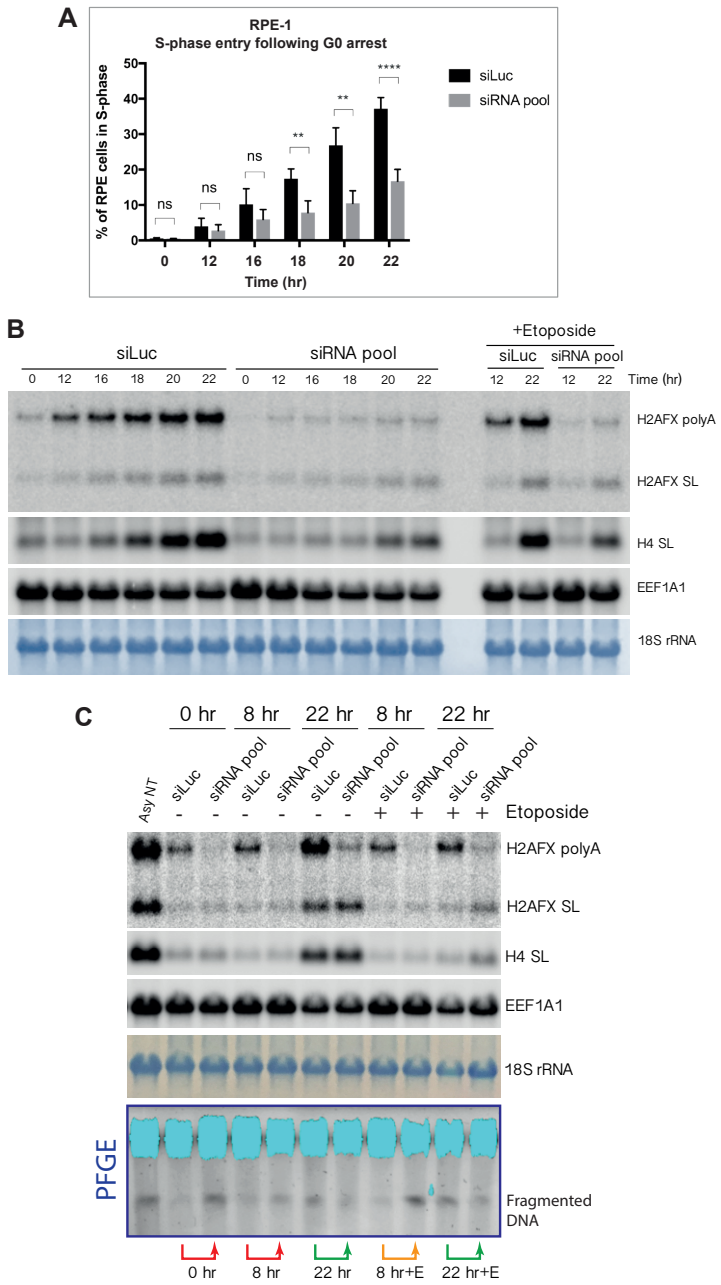



**Figure 4.2 H2A.X poly(A)$^+$ mRNA depletion in RPE-1 cells.** The H2A.X poly(A)$^+$ mRNA "knock down" was performed using either a pool of 4 siRNAs (siRNA pool) or a single siRNA targeting the extended 3'-UTR (single siRNA), as represented in Figure 4.1A. A) Percentage of cells in S-phase in RPE-1 cells treated with either siLuc (black bars) or siRNA pool (gray bars) and synchronised by contact inhibition. The x-axis represents the time points in hr after the release from the cell cycle arrest and the y-axis the percentage of S-phase cells determined by EdU labelling followed by flow cytometry (Supplementary Figure S5). The error bars represent the standard error of four independent experiments calculated by multiple t-tests using the Graphpad Prism software. ns: not significant. ** P-value $\leq$0.01. **** P-value $\leq$ 0.0001. B) Total RNA northern blot of the samples shown in E as well as samples treated with 5 µM etoposide for 2 hr before harvesting (last four lanes). C) Total RNA northern blot of RPE-1 cells treated with the corresponding siRNAs and arrested for 4 days and harvested at the indicated time points in hr after resuming the cell cycle. 5 µM etoposide were applied at 6 hr (G1) for a 2 hr period. The same samples were analysed by pulsed-field gel electrophoresis (PFGE). The dark blue box represents the ethidium bromide-stained PFGE agarose gel. The light blue regions correspond to saturated DNA signal in the well of the gel. H2A.X SL mRNA and H2A.X poly(A)$^+$ mRNA were detected with a probe complementary to the common region between both mRNA isoforms. H4 SL: canonical H4 SL mRNA. EEF1A1: eukaryotic translation elongation factor 1 alpha 1. 18S rRNA: 18S ribosomal RNA (methylene blue staining).

## 4.2 Effect of H2A.X poly(A)$^+$ mRNA depletion on DNA damage signalling

Experiments were next performed to assess the role of H2A.X poly(A)$^+$ mRNA under DNA damage conditions. In particular, H2A.X poly(A)$^+$ mRNA depleted HeLa cells were treated with etoposide for 2 hr before harvesting and H2A.X as well as $\gamma$H2A.X protein levels were analysed by western blotting (Figure 4.3A). The topoisomerase II inhibitor etoposide was the DNA damaging agent of choice as it induces DNA damage in all the phases of the cell cycle [320]. Compared to untreated cells, treatment with etoposide leads to the accumulation of H2A.X protein in control cells (siLuc) but also in H2A.X poly(A)$^+$ mRNA depleted cells, although to a lesser extent (Figure 4.3A). This effect is likely to be due to a change in H2A.X mRNA translation or protein stability as H2A.X poly(A)$^+$ mRNA levels are unaffected by the DNA damage inducing agent (Figure 4.3B), as also observed in RPE-1 cells (Figure 4.2B). As previously shown [321], etoposide treatment led to the accumulation of $\gamma$H2A.X. H2A.X poly(A)$^+$ mRNA depleted cells accumulate lower $\gamma$H2A.X levels compared to the siLuc control, correlating to the total H2A.X protein levels in these samples (Figure 4.3A, lower panel).

As S-phase cells still express the H2A.X SL mRNA, the remaining $\gamma$H2A.X protein levels observed in H2A.X poly(A)$^+$ mRNA depleted cells by western blot may reflect the $\gamma$H2A.X levels produced from the H2A.X SL mRNA isoform in S-phase cells. Thus to distinguish between the $\gamma$H2A.X signal in S-phase and non-S-phase HeLa cells, the H2A.X poly(A)$^+$ mRNA depleted cells were labelled with EdU, which marks S-phase cells, and stained with DAPI and an antibody specific for $\gamma$H2A.X by immunofluorescence (Figure 4.3C and D). As expected, only etoposide-treated cells showed $\gamma$H2A.X (compare panels C and D of Figure 4.3). The H2A.X poly(A)$^+$ mRNA depletion led to a significant general decrease in $\gamma$H2A.X signal, in both non-S-phase but surprisingly also S-phase cells (Figure 4.3E).

To assess whether the observed phenotype is specifically caused by the depletion of the H2A.X poly(A)$^+$ mRNA, an siRNA resistant wild-type H2AFX gene (deleted "single siRNA" binding site or "dsi") was stably integrated into a HeLa Flp-In T-REx cell line. As an additional control, an siRNA resistant PAS deletion mutant (AAUAAA deleted) was

assessed in parallel. An illustration of the pre-mRNA generated by these two transgenes is shown in Figure 4.4A (see Chapter 2, Subsection 2.5.5 for the details on how these constructs were made). These cells were transfected with siLuc, the siRNA pool or the single siRNA for 72 hr. Equal amounts of total RNA were analysed by northern blotting. As expected, both these cell lines were sensitive to the siRNA pool (Figure 4.4B). However, only the wild-type siRNA-resistant transgene was able to accumulate H2A.X poly(A)$^+$ mRNA after the treament with the single siRNA (Figure 4.4B, compare H2AFX-HindIII-dsi and H2AFX-HindIII-dsi-dPAS). This result suggests that the PAS deletion abolishes processing by cleavage and polyadenylation and thus the production of siRNA-resistant transgene H2A.X poly(A)$^+$ mRNA. The levels of H2A.X poly(A)$^+$ mRNA in siLuc-treated cells represent the sum of the endogenous and the transgene H2A.X poly(A)$^+$ mRNA. Unlike the siRNA-resistant transgene H2A.X poly(A)$^+$ mRNA, the endogenous H2A.X poly(A)$^+$ mRNA is sensitive to the single siRNA, thus H2AFX-HindIII-dsi cells treated with the single siRNA show decreased H2A.X poly(A)$^+$ mRNA levels compared to siLuc. In contrast, cells expressing the H2AFX PAS deletion transgene do not express H2A.X poly(A)$^+$ mRNA and thus the H2A.X poly(A)$^+$ mRNA levels detected in the siLuc sample exclusively correspond to endogenous H2A.X poly(A)$^+$ mRNA (Figure 4.4B).

The etoposide-treated samples highlighted in red boxes in Figure 4.4B were used for immunofluorescence to assess $\gamma$H2A.X formation upon DNA damage in these cell lines (Figure 4.5). Correlating with the northern blot results, the siRNA pool reduces the $\gamma$H2A.X signal in both these cell lines, independently of the cell cycle phase, as observed wild-type HeLa cells (Figure 4.3D). However, the $\gamma$H2A.X signal observed in cells expressing the wild-type siRNA-resistant H2AFX transgene (H2AFX-HindIII-dsi) transfected with the single siRNA was comparable to the siLuc control (Figure 4.5A). In contrast, the single siRNA-treatment did not allow $\gamma$H2A.X formation in the cells containing the siRNA-resistant H2AFX PAS deletion mutant (H2AFX-HindIII-dsi-dPAS) that does not allow transgene H2A.X poly(A)$^+$ mRNA expression (Figure 4.5B). The fact that the phenotype can be rescued by the insertion of a wild-type siRNA resistant H2AFX gene, proves that the reduction in $\gamma$H2A.X signal is exclusively due to the absence of the H2A.X poly(A)$^+$ mRNA. Thus, the H2A.X poly(A)$^+$ mRNA is required for efficient DNA damage signalling in both non-S-phase and S-phase cells.

It has been previously shown that in resting cells, DNA damage induction leads to the stabilisation of H2A.X protein and *de novo* H2A.X incorporation into the chromatin to allow efficient γH2A.X foci formation [322]. In HeLa cells, H2A.X protein also increases upon the treatment with DNA damaging agent (Figure 4.3A). This could be due to a similar protein stabilisation effect as observed in resting cells or to stimulated translation. In H2A.X poly(A)$^+$ mRNA depleted cells, the only *de novo* H2A.X protein source is the H2A.X SL mRNA. However, S-phase cells also show reduced γH2A.X signal when the H2A.X poly(A)$^+$ mRNA is depleted. This begs the question as to what happens to S-phase-specific H2A.X SL mRNA under DNA damage conditions. Replication stress is known to promote degradation of histone SL mRNA [323]. Therefore, to test whether the H2A.X SL mRNA is degraded as a consequence of the etoposide treatment, H2A.X SL mRNA levels were monitored in HeLa cells treated with etoposide over a time period of 5 hr. As expected, etoposide treatment induced DNA damage as shown by the accumulation of γH2A.X (Figure 4.6A). Consistent with a previous report [256], DNA damage induction leads to partial reduction of RD histone SL mRNA levels and to a lesser extent to reduced H2A.X SL mRNA levels. However, five times more etoposide (25 µM instead of 5 µM) were required to observe this effect and yet a substantial amount of H2A.X SL mRNA is detected at the 2 hr time point (time point used in immunofluorescence experiments) (Figure 4.6B). Thus, the H2A.X SL mRNA is partially degraded but still largely present after 2 hr etoposide treatment.

Another possibility is that following DNA damage H2A.X SL mRNA is retained in the nucleus and so avoids cytoplasmic translation. To test this hypothesis, HeLa cells were treated with 5 µM etoposide (same concentration as used in immunofluorescence experiments) for 1 to 5 hr and separated into nuclear and cytoplasmic fractions. Fractionation efficiency was confirmed by western blotting using antibodies against tubulin (exclusively cytoplasmic) and fibrillarin (exclusively nuclear) (Figure 4.6C). H2A.X poly(A)$^+$ mRNA is relatively more abundant in the nucleus compared to H2A.X SL mRNA but its levels are not affected by the etoposide treament (Figure 4.6D). H2A.X poly(A)$^+$ mRNA export may be enhanced upon etoposide treatment as it appears to slightly increase in the cytoplasmic fractions, however more experiments are required to test such an effect. In contrast, H2A.X SL and H4 SL mRNA levels gradually diminished in both nuclear and cytoplasmic compartments over the 5 hr timecourse. This effect is less pronounced for H2A.X SL

mRNA and at 2 hr a substantial amount is still present in the cytoplasm. Taken together, the H2A.X SL mRNA is only partially degraded upon 5 µM etoposide treatment and is still largely present in the cytoplasm.

A further possibility is that the H2A.X SL mRNA is not translated under the tested conditions. Polysome gradients followed by northern blotting were performed on untreated or HeLa cells treated with 5 µM for 2 hr to test this possibility. H2A.X SL and H4 SL mRNA is still detected on polysomes in the etoposide-treated cells (Figure 4.7). However, the H2A.X SL/poly(A)$^+$ mRNA ratio slightly changed upon DNA damage induction, with a higher proportion of H2A.X poly(A)$^+$ mRNA in the polysome fractions (Figure 4.7, see poly(A)/SL ratio in fractions 8 and 9). Nevertheless, it is unclear whether this effect is due to decreased H2A.X SL mRNA loading onto polysomes or increased loading of H2A.X poly(A)$^+$ mRNA. The former is possible as H4 SL levels are slightly decreased. Even so it appears that the effects on translation are only mild.

The above listed explanations assume that *de novo* H2A.X is required for efficient $\gamma$H2A.X foci formation upon DNA damage, as has been suggested [322]. However, the decreased $\gamma$H2A.X signalling in H2A.X poly(A)$^+$ mRNA depleted cells could also represent lower H2A.X incorporation into the chromatin in S-phase. The H2A.X poly(A)$^+$ mRNA levels in S-phase HeLa cells are relatively high compared to its SL counterpart. Therefore, it is possible that a 72 hr depletion, which involves at least two rounds of DNA replication in the absence of the H2A.X poly(A)$^+$ mRNA (depleted at 24 hr of siRNA treatment) leads to less H2A.X deposition in the newly synthesised chromatin and thus to decreased $\gamma$H2A.X signalling upon DNA damage in these cells.

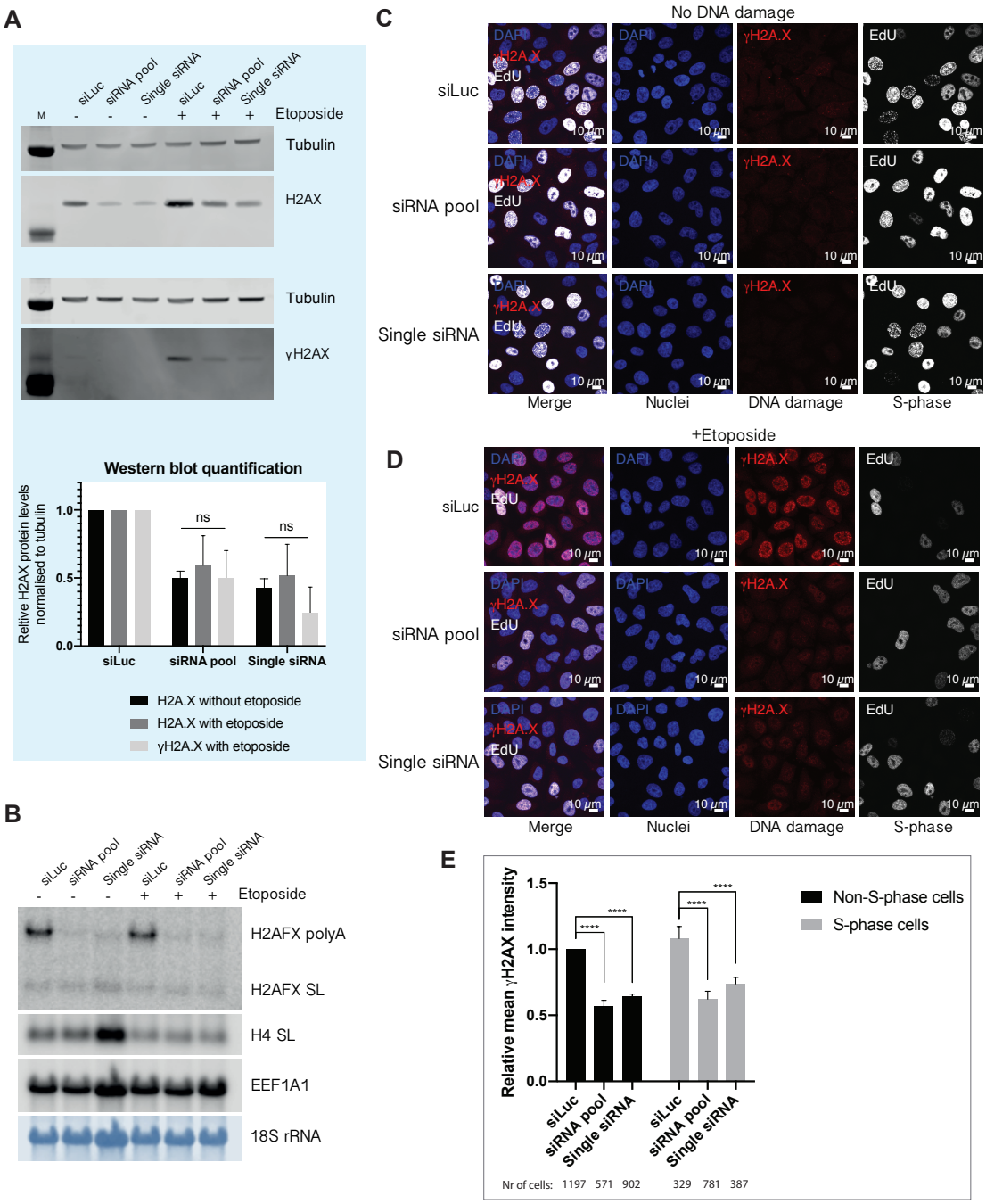



**Figure 4.3 H2A.X poly(A)$^+$ mRNA depletion leads to decreased γH2A.X signalling.** A) Total protein extract western blot of cells treated with siRNA for 72 hr (representative blot of three independent experiments). The etoposide treatment (5 µM) was performed for 2 hr before harvesting. Tubulin was used as loading control. H2AX: total H2A.X protein. γH2A.X: H2A.X phosphorylated at Ser139. The bottom graph shows the western blot quantification of three independent experiments. The error bars represent the standard error calculated by two-way Anova performed using the Graphpad Prism software. ns: p-value >0.05. B) Total RNA northern blot of the samples shown in A (representative blot of three independent experiments). C) Immunofluorescence of the cells shown in A without etoposide treatment. Images were processed in the ImageJ software. D) With 5 µM etoposide treatment for 2 hr. Blue: DAPI (nuclei). Red: γH2A.X (DNA damage). White: EdU (S-phase cells). E) Quantification of nuclear γH2A.X signal in non-S-phase (black bars) and S-phase cells (gray bars) of the experiment shown in D, using a macro in ImageJ to ensure unbiased quantification. The x-axis indicates the used siRNAs and the y-axis the relative mean γH2A.X maximum intensity (compared to siLuc in non-S-phase cells). For the quantification, 10 images of each sample were analysed for a total of three independent experiments. The total number (Nr) of cells analysed is indicated below the graph. The error bars represent the standard error calculated by two-way Anova performed using the Graphpad Prism software. **** P-value $\leq$ 0.0001. M: protein size marker.

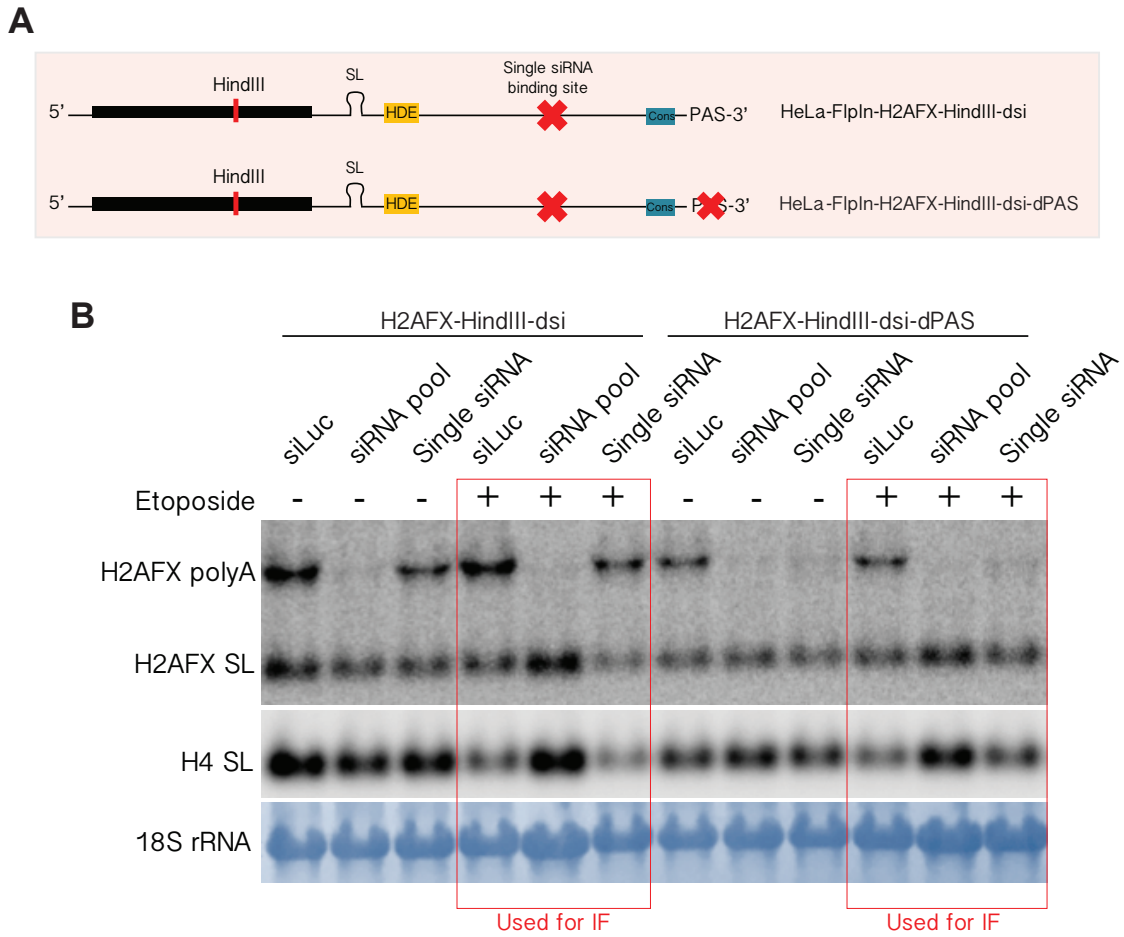



**Figure 4.4 H2A.X poly(A)$^+$ mRNA depletion rescue experiment.** A) HeLa Flp-In T-REx cells expressing an siRNA resistant (deleted "single siRNA" binding site) copy of either wild-type H2AFX (HeLa-FlpIn-H2AFX-HindIII-dsi) or an H2AFX PAS deletion mutant (deleted AAUAAA, HeLa-FlpIn-H2AFX-HindIII-dsi-dPAS) were transfected with siLuc, siRNA pool or single siRNA for 72 hr and treated with 5 µM etoposide for 2 hr. B) Total RNA was used for northern blotting. H2A.X SL mRNA and H2A.X poly(A)$^+$ mRNA isoforms were detected with a probe complementary to the common region between both mRNA isoforms. H4 SL: canonical H4 SL mRNA. 18S rRNA: 18S ribosomal RNA (methylene blue staining). The red boxes highlight the samples used for immunofluorescence, shown in Figure 4.5. The H2AFX transgene contains a HindIII site in the ORF to be distinguished from the endogenous by PCR (see Chapter 6).

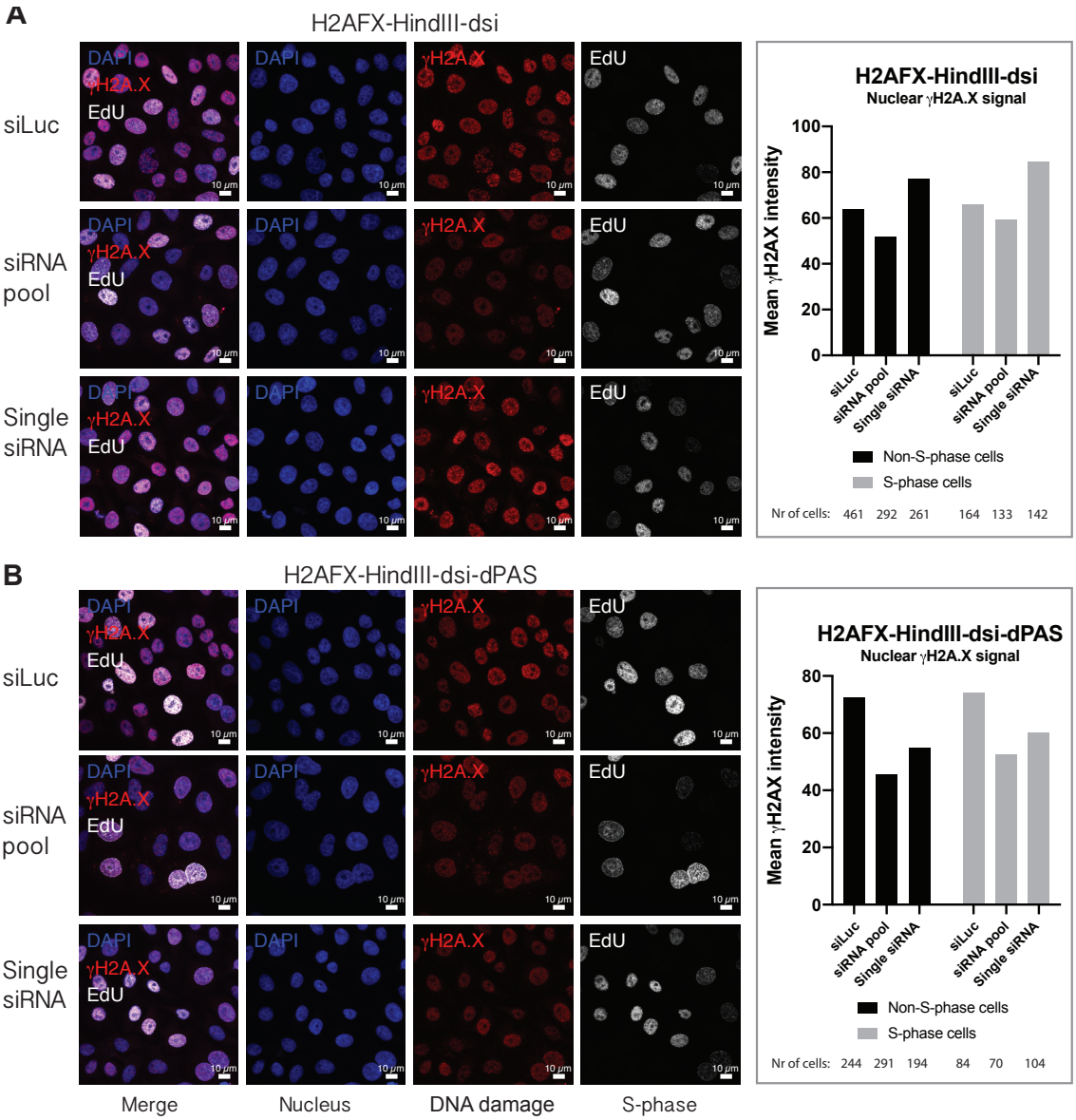



**Figure 4.5 Only siRNA resistant wild-type H2AFX rescues the H2A.X poly(A)$^+$ mRNA-depletion phenotype.** Immunofluorescence images of the samples highlighted in red boxes in Figure 4.4. A) siRNA-resistant wild-type H2AFX (H2AFX-dsi). B) siRNA-resistant H2AFX PAS deletion mutant (deleted AAUAAA, H2AFX-dsi-dPAS). The cells were labelled with EdU (S-phase cells, white) and stained with DAPI (nuclei, blue) and γH2A.X antibody by immunofluorescence (DNA damage, red). Images were processed in the ImageJ software. For the quantification (panels on the right), 10 images of each sample were analysed. The total number (Nr) of cells analysed is indicated below the graph.

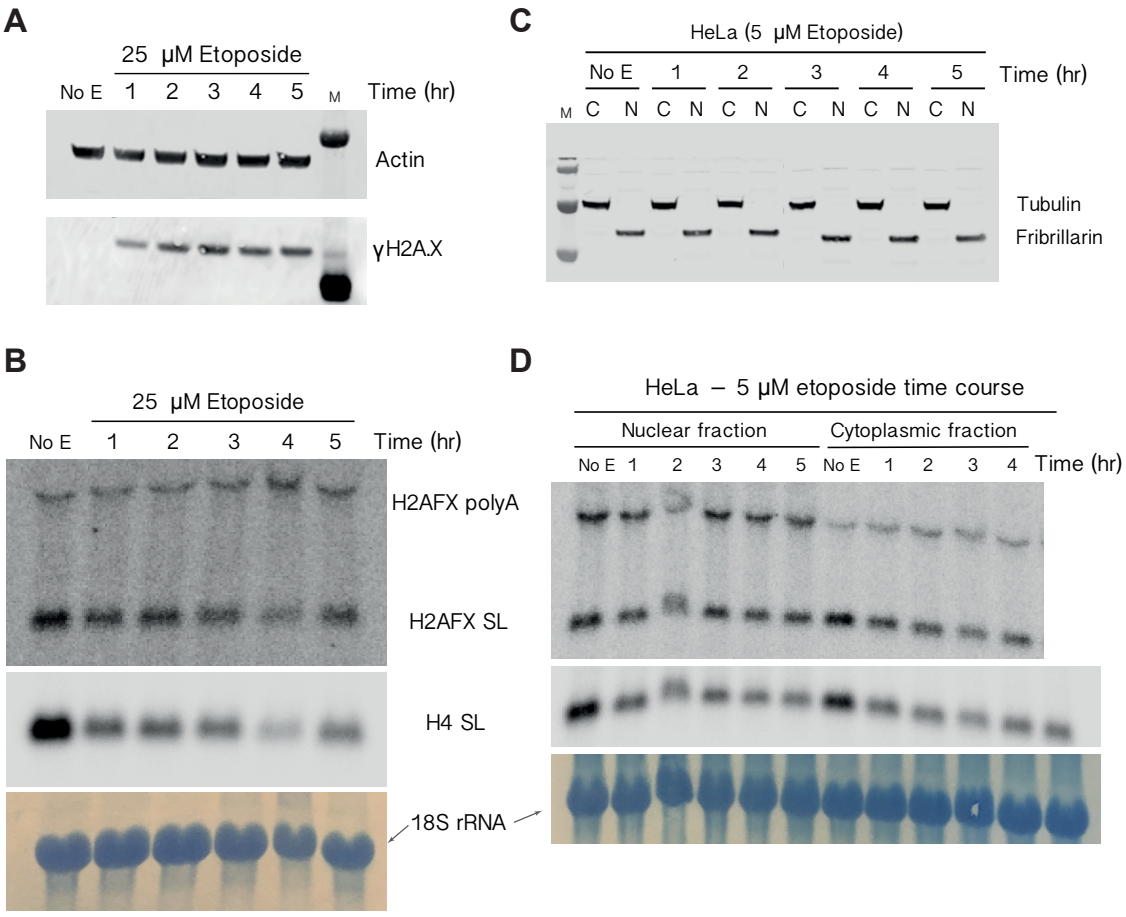



**Figure 4.6 H2A.X SL mRNA is partially decreased but still largely present in nucleus and cytoplasm following etoposide treatment.** HeLa cells were treated with etoposide for 1, 2, 3, 4 or 5 hr. A) Western blot of total protein extracts showing etoposide-induced γH2A.X accumulation. B) Total RNA northern blot of the samples shown in A. C) Western blot of nuclear and cytoplasmic fractions using antibodies against tubulin (exclusively cytoplasmic) and fibrillarin (exclusively nuclear). D) Northern blot of the samples shown in C. H2A.X SL mRNA and H2A.X poly(A)$^+$ mRNA isoforms were detected with a probe complementary to the common region between both mRNA isoforms. H4 SL: canonical H4 SL mRNA. 18S rRNA: 18S ribosomal RNA (methylene blue staining). No E: not treated with etoposide. M: protein size marker.

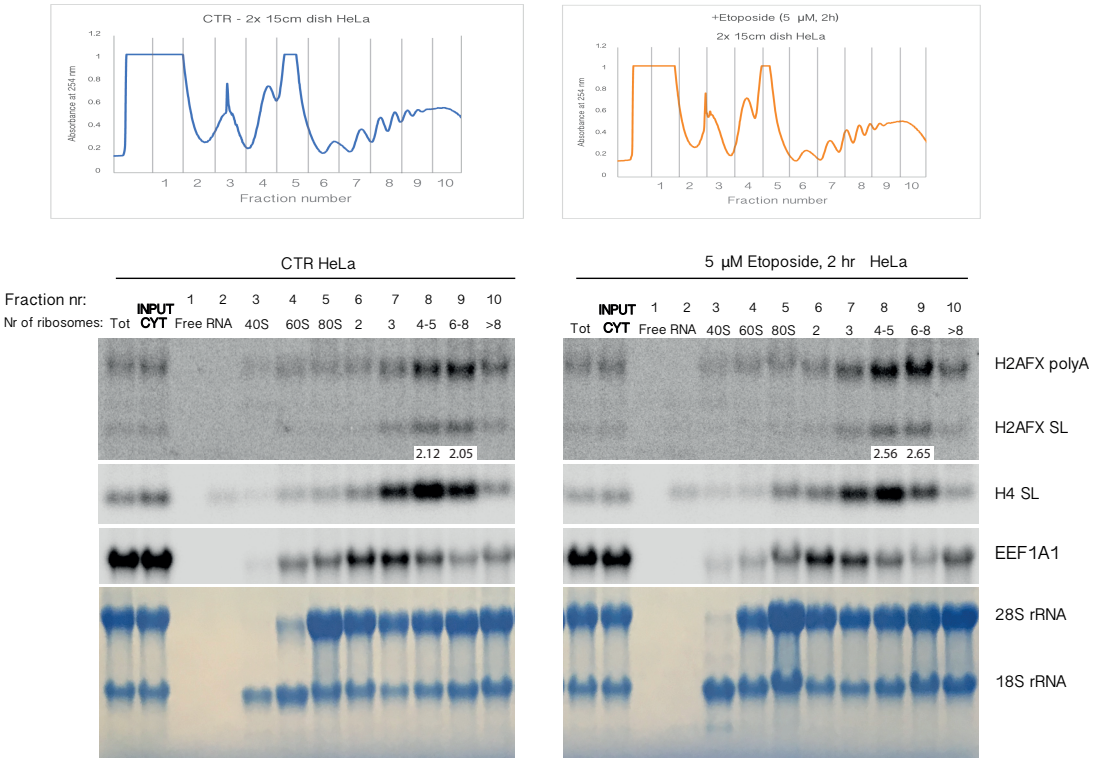



**Figure 4.7 DNA damage caused by 5 µM etoposide mildly affects H2A.X mRNA translation.** Cytoplasmic lysates of HeLa cells untreated (top left panel) or treated with 5 µM etoposide (top right panel) were subjected to polysome gradient fractionation. RNA was extracted from each fraction and the same RNA volume was loaded per lane of the northern blot gel. The fraction number and the corresponding number of ribosomes bound to the RNA are indicated above each lane. H2A.X SL mRNA and H2A.X poly(A)$^+$ mRNA isoforms were detected with a probe complementary to the common region between both mRNA isoforms. H4 SL: canonical H4 SL mRNA. EEF1A1: eukaryotic translation elongation factor 1 alpha 1. 18S rRNA: 18S ribosomal RNA (methylene blue staining). The numbers below the H2A.X mRNA bands for fractions 8 and 9 correspond to the H2A.X poly(A)$^+$/SL RNA ratio calculated by densitometry using the ImageStudioLite software.

# Chapter 5

# H2A.X SL mRNA

## 5.1 H2A.X SL mRNA "knock out"

The sequence of the H2A.X SL mRNA is shared with its extended polyadenylated counterpart, thus the H2A.X SL mRNA does not contain unique sequences that can be specifically targeted by RNAi. Therefore, the H2A.X SL mRNA can only be depleted by preventing U7-dependent processing through targeted mutation of the histone downstream element (HDE). Notably, deletion of 9 nucleotides of the HDE is known to abolish U7-dependent processing of histone mRNA in mammalian cells [324]. To confirm this for H2AFX, five expression plasmids were constructed, expressing the EGFP ORF followed by different 3'-UTR sequences (Figure 5.1, top left "Plasmid constructs"): 1) a short 3'-UTR containing the bovine growth hormone (bGH) polyadenylation site (PAS), 2) the extended wild-type H2AFX 3'-UTR, 3) the extended H2AFX 3'-UTR with the 9-nucleotide deletion of the HDE, 4) a short wild-type H2AFX 3'-UTR until past the HDE, and 5) a short wild-type H2AFX 3'-UTR until past the HDE but with the 9-nucleotide HDE deletion. These constructs were transiently transfected into HeLa cells for 24 hr and the resulting RNA products were analysed by northern blotting using a probe targeting the EGFP ORF (Figure 5.1).

The cells transfected with the construct only containing the EGFP ORF followed by the bGH PAS (#1) only expressed one RNA isoform (calculated as 928 ribonucleotides in length plus the lenth of the poly(A)-tail). In contrast, construct #2, containing the wild-type H2A.X 3'-UTR, expressed both a short and a long RNA isoform. These result

from U7-dependent processing at the H2A.X SL (835 ribonucleotide RNA) and cleavage and polyadenylation (CPA) at the downstream H2A.X PAS (1872 ribonucleotide RNA plus the lenth of the poly(A)-tail), respectively. The same construct with the HDE deletion (#3) only expressed the long RNA isoform, indicating U7-dependent processing was abolished. Similarly, the construct containing the short wild-type H2A.X 3'-UTR (#4) expressed both short (processed at the U7-dependent site, 829 ribonucleotides) and long RNA isoforms (processed at the bGH PAS, 1048 ribonucleotides plus the lenth of the poly(A)-tail), whereas its HDE deletion counterpart (#5) expressed only the longer RNA. These results confirm that a 9-nucleotide HDE deletion indeed abolishes H2AFX U7-depending processing and thus the production of H2A.X SL mRNA (Figure 5.1).

Encouraged by the above results, a 24 nt region containing the 9 bp region of the H2AFX HDE was targeted in both HeLa and HCT116 cells for deletion by CRISPR-Cas9 technology. One homozygous clone was obtained for each cell line (referred to as HeLa dHDE and HCT116 dHDE, respectively). The only homozygous HeLa dHDE clone also contains a 90 bp deletion downstream of the HDE deletion (this was the only sequencing product obtained, see Figure 5.2). This deleted region does not contain any known sequence elements involved in the regulation of the H2A.X poly(A)$^+$ mRNA and so was accepted for futher study. Importantly, the SL structure is intact in both HeLa and HCT116 dHDE clones. Consequently, the H2A.X poly(A)$^+$ mRNA expressed in these cell lines still contains the SL.

As shown by northern blotting, the HDE deletion efficiently abolishes processing at the H2AFX U7-dependent processing site (Figure 5.3A). Interestingly, H2A.X poly(A)$^+$ mRNA levels are unchanged in dHDE cells comparing to wild-type cells (Figure 5.3A, right panel). Furthermore, compared to HeLa cells, HCT116 cells express much more SL than poly(A)$^+$ H2A.X mRNA. Even though, asynchronously growing wild-type HCT116 cells show 10% more cells in S-phase compared to wild-type HeLa cells, this does not explain the difference in the H2A.X SL mRNA levels because H4 SL mRNA levels are not increased in HCT116 cells compared to HeLa cells (Figure 5.3A and B).

Both HeLa and HCT116 dHDE clones show a minor reduction in S-phase cell number (5-10%) compared to their respective wild-type counterpart (Figure 5.3B). Interestingly, the deletion of the HDE influences cell morphology in a different manner in the two cell

lines, HeLa dHDE cells show longer while HCT116 dHDE cells show shorter protrusions (Figure 5.3C).

Taken together, the H2AFX HDE deletion abolishes U7-dependent processing, generating a complete H2A.X SL mRNA "knock out" in both cell types. Notably, asynchronous dHDE cell populations show a mild reduction in S-phase cells and cell line-specific morphological alterations.

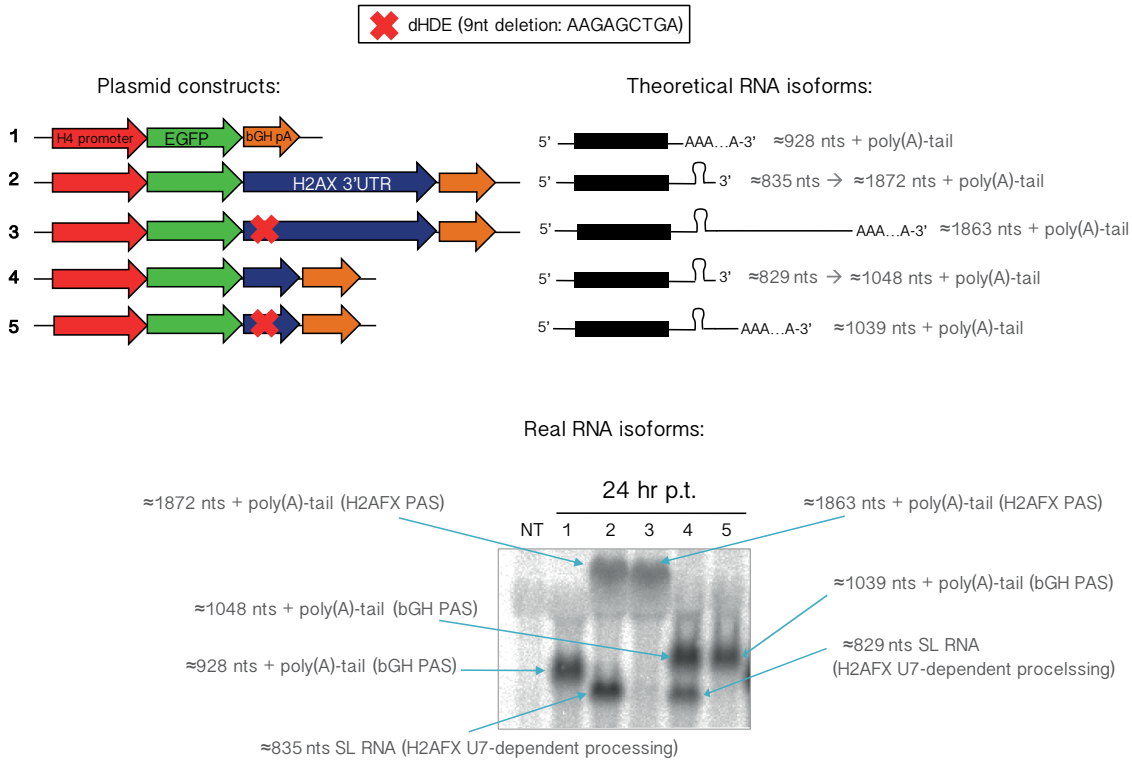



**Figure 5.1 A 9-nucleotide deletion of the HDE abolishes H2AFX U7-dependent processing.** Five expression plasmids were constructed, containing the EGFP ORF under the control of the histone H4 gene promoter and followed by different 3'-UTR sequences: 1) a short 3'-UTR containing the bGH PAS, 2) the extended wild-type H2AFX 3'-UTR, 3) the extended H2AFX 3'-UTR with the 9-nucleotide deletion of the HDE, 4) a short wild-type H2AFX 3'-UTR until past the HDE, and 5) a short wild-type H2AFX 3'-UTR until past the HDE but with the 9-nucleotide deletion of the HDE (see "Plasmid constructs" on the left). The "Theoretical RNA isoforms" generated by each construct are represented on the right. These constructs were transiently transfected into HeLa cells for 24 hr (24h p.t.) and the "Real RNA isoforms" were analysed by northern blotting using a probe complementary to the EGFP ORF (bottom). 24 hr p.t.: 24 hr post transfection. NT: not transfected. nts: ribonucleotides. 28S and 18S rRNA bands from the methylene blue staining of the membrane were used as size references (data not shown). The indicated sizes are the theoretical sizes inferred from the plasmid constructs.

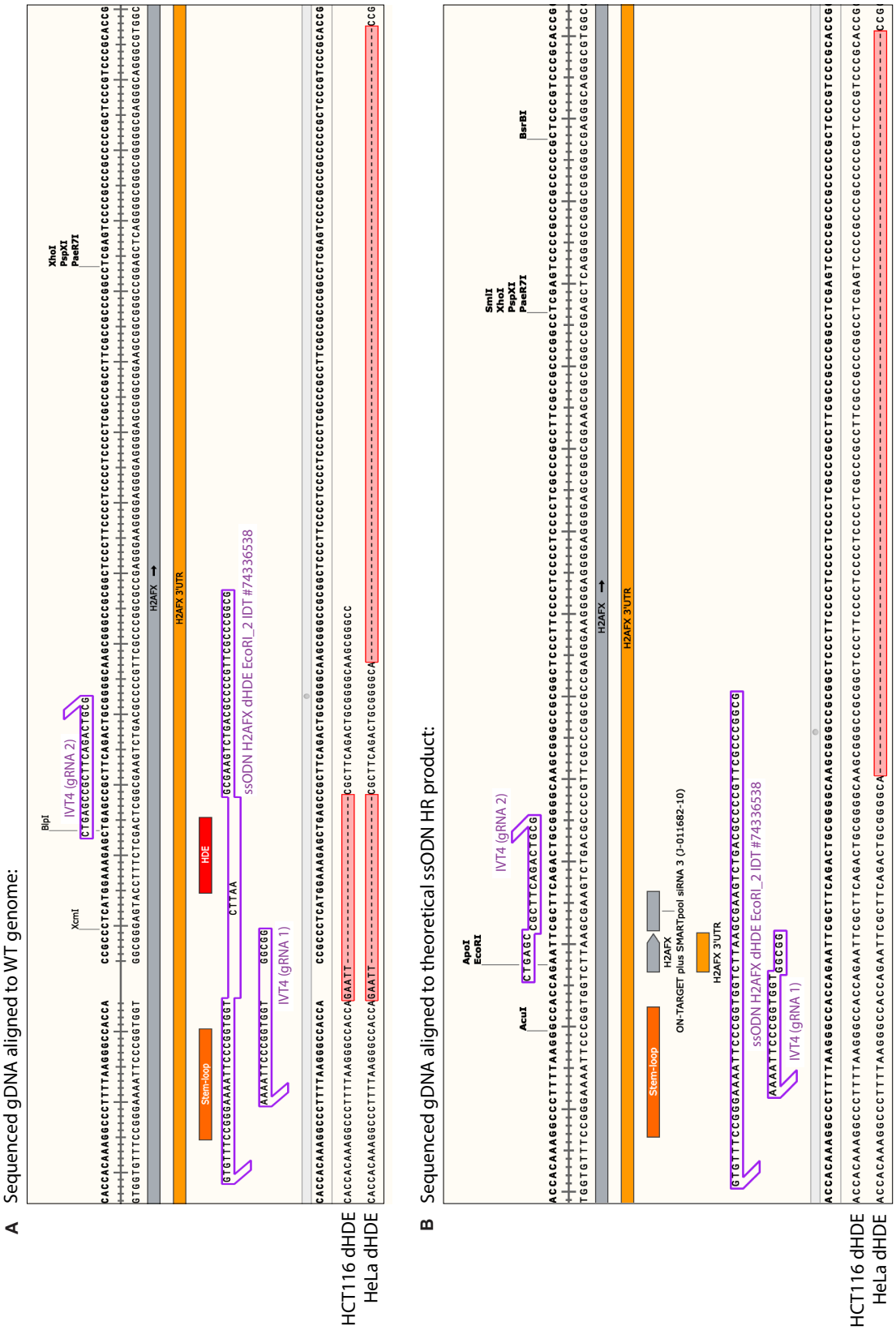



**Figure 5.2 Genomic DNA sequences of HeLa and HCT116 CRISPR dHDE clones.** The gDNA of HeLa and HCT116 CRISPR dHDE clones was amplified by PCR and sequenced. A) Aligned to the human wild-type H2AFX gene. Note the EcoRI site (GAATTC) inserted by the ssODN at the location of the HDE deletion. B) Aligned to the theoretical sequence created by homologous recombination (HR) using the ssODN as a template. Note the 90 bp deletion downstream of the HDE in the HeLa dHDE clone. IVT4 and 5 show the two gRNAs (gRNA 1 and gRNA 2, respectively, from Table 2.7). ssODN: single-stranded donor oligonucleotide. HDE: histone downstream element. These images were obtained using SnapGene software.

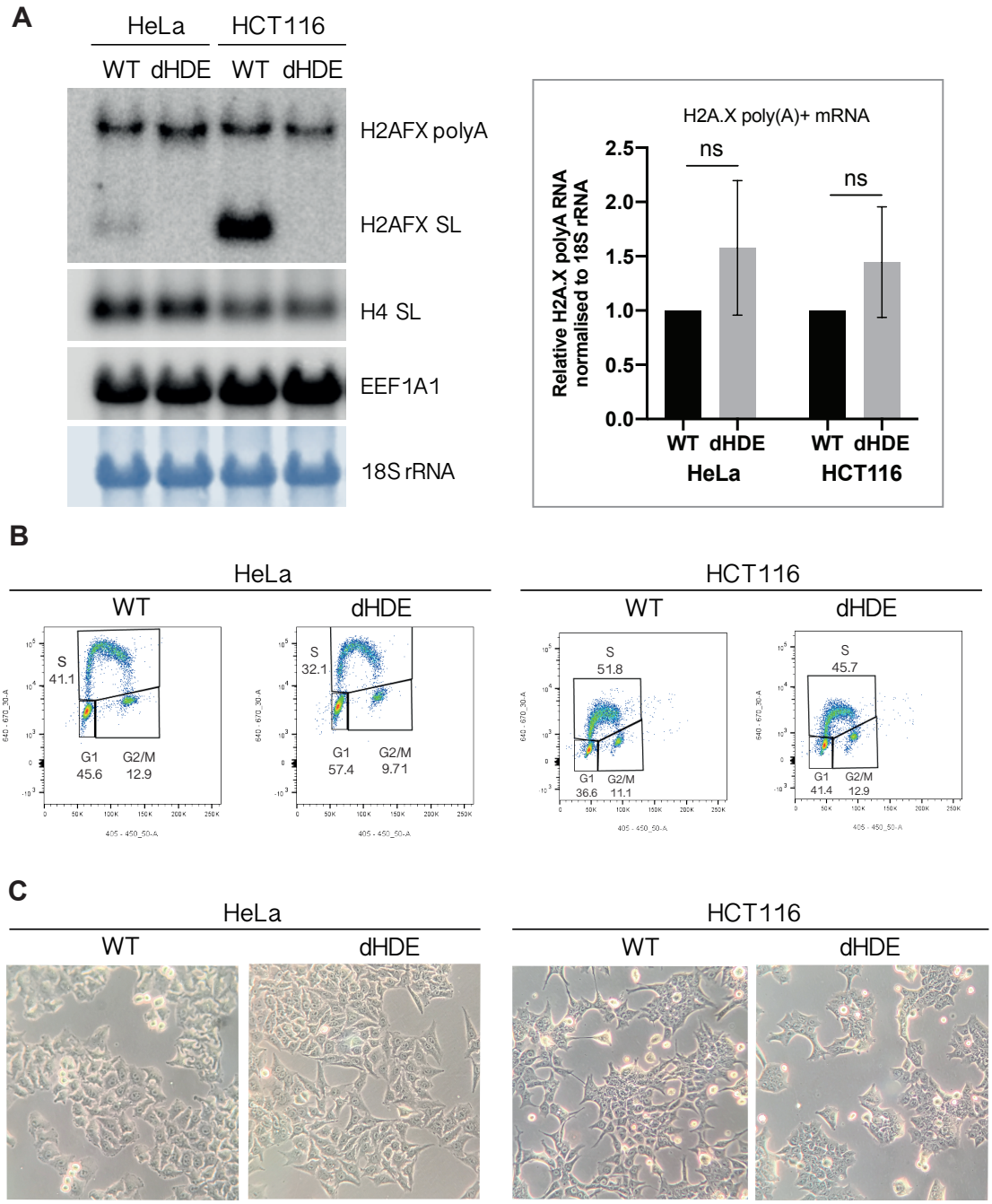



**Figure 5.3 The H2AFX HDE deletion efficiently impedes H2A.X SL mRNA expression.** A) Total RNA northern blot of asynchronously growing wild-type (WT) and dHDE HeLa and HCT116 cells (representative northern blot of three independent experiments). H2A.X SL mRNA and H2A.X poly(A)$^+$ mRNA isoforms were detected with a probe complementary to the common region between both mRNA isoforms. H4 SL: canonical H4 SL mRNA. EEF1A1: eukaryotic translation elongation factor 1 alpha 1. 18S rRNA: 18S ribosomal RNA (methylene blue staining). The graph on the right shows the H2A.X poly(A)$^+$ mRNA levels in WT and dHDE cells, calculated by densitometry from 6 independent experiments. Significance was calculated using a paired t-test. ns: p-value $> 0.05$. B) Flow cytometry of EdU-labelled cells. The graphs show the DAPI fluorescent intensity on the x-axis and Alexa 647-EdU fluorescent intensity on the y-axis. Percentages of cells in G1, S and G2/M were determined using the FlowJo software. C) Bright-field image showing cell morphology.

## 5.2 Effect of H2A.X SL mRNA "knock out" on the cell cycle distribution of H2A.X poly(A)$^+$ mRNA

In asynchronously growing dHDE cells (*i.e.* H2A.X SL mRNA "knock out" cells) the levels of H2A.X poly(A)$^+$ mRNA does not show a significant change (Figure 5.3A, right panel). This was surprising as the HDE deletion might be anticipated to increase the level of read-through transcription up to the PAS. To test whether the H2A.X poly(A)$^+$ mRNA cell cycle distribution is altered when U7-dependent processing is abolished, dHDE cells were synchronised by DTB and time points were collected over a period of 16 hr following DTB release (see Supplementary Figures S6 and S7 for flow cytometry data). As shown by northern blotting (Figure 5.4), the H2A.X poly(A)$^+$ mRNA in both HeLa and HCT116 dHDE cells is only mildly increased in S-phase (Figure 5.4A and B, respectively), suggesting that there is an RNA stability mechanism limiting the generation of H2A.X poly(A)$^+$ mRNA in S-phase, potentially by degrading the pre-mRNA that has not been processed at the U7-dependent cleavage site. In contrast, in G2/M the levels of H2A.X poly(A)$^+$ mRNA of dHDE cells markedly increased. The H2A.X poly(A)$^+$ mRNA in dHDE cells contains the intact SL structure. Since SLBP is known to bind to the SL of H2A.X poly(A)$^+$ mRNA [174], it is conceivable that SLBP plays a role in controlling H2A.X poly(A)$^+$ mRNA generation. A possible mechanism for such regulation is discussed in Chapter 7.

Interestingly, wild-type HCT116 cells express much higher levels of H2A.X SL mRNA in S-phase compared to HeLa cells (Figure 5.4). This demonstrates that the higher H2A.X SL mRNA levels observed in asynchronously growing cells (Figure 5.3A) are not only caused by the higher percentage in S-phase cells but mainly by increased accumulation of H2A.X SL mRNA in S-phase in HCT116 cells.

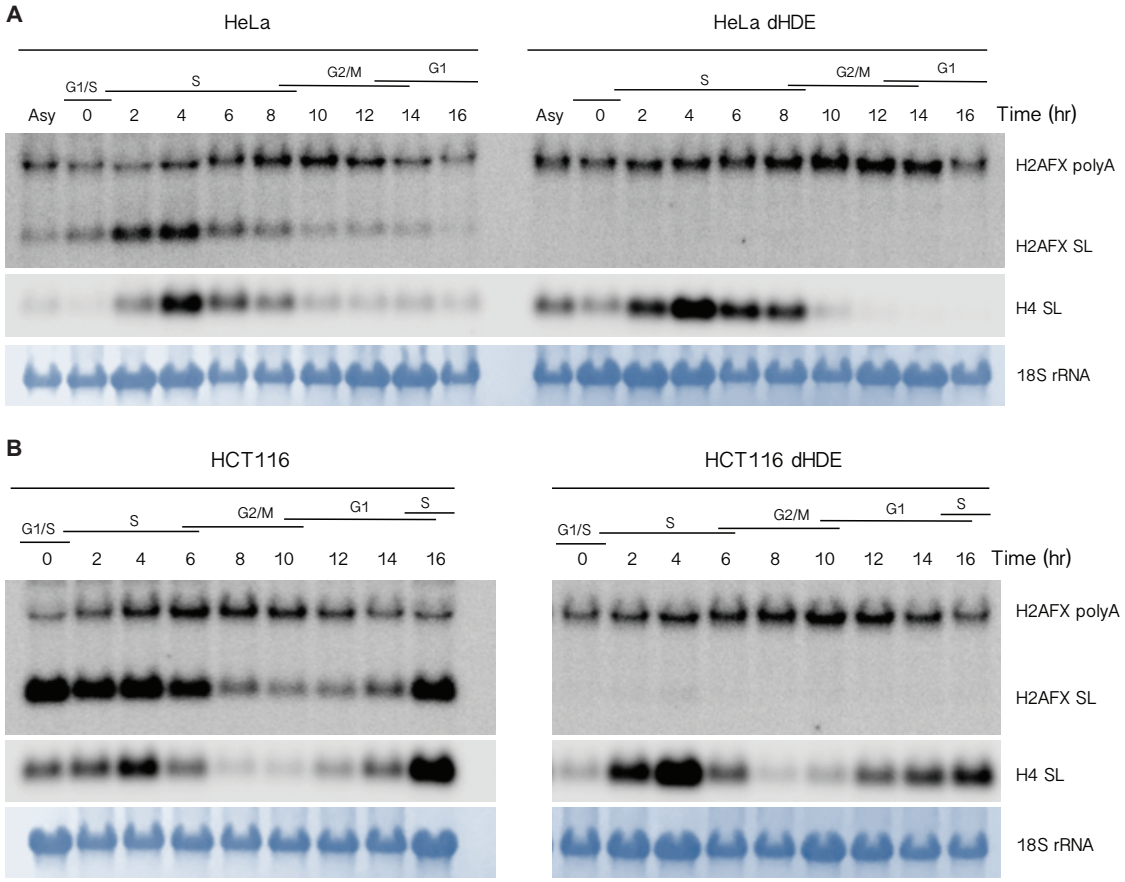



**Figure 5.4 Cell cycle distribution of the H2A.X poly(A)$^+$ mRNA in H2A.X SL mRNA "knock out" HeLa and HCT116 cells (dHDE).** Wild-type and dHDE cells were synchronised by DTB and total RNA was analysed by northern blotting: A) HeLa, B) HCT116 cells. The time points indicate the time in hr after the release from the DTB. The corresponding cell cycle phases determined by flow cytometry (Supplementary Figures S6 and S7) are indicated above. H2A.X SL mRNA and H2A.X poly(A)$^+$ mRNA isoforms were detected with a probe complementary to the common region between both mRNA isoforms. H4 SL: canonical H4 SL mRNA. 18S rRNA: 18S ribosomal RNA (methylene blue staining). Asy: asynchronously growing cells. This experiment was performed only once, repetitions are required to calculate statistics.

## 5.3 Effect of H2A.X SL mRNA "knock out" on DNA damage signalling

As shown in Chapter 4, H2A.X poly(A)$^+$ mRNA depletion has a significant effect on DNA damage signalling in HeLa cells. To assess whether H2A.X SL mRNA "knock out" cells show a similar effect, wild-type and dHDE cells were treated with 5 µM etoposide for 2 hr and H2A.X RNA and protein levels were then analysed.

Notably, the etoposide treatment has no significant effect on H2A.X poly(A)$^+$ mRNA levels in both wild-type and dHDE HeLa and HCT116 cells (Figure 5.5, bottom panels). Interestingly, H2A.X SL and H4 SL mRNA levels drastically decrease upon etoposide treatment in HCT116 cells but not in HeLa cells (at least 25 µM etoposide are required to see this effect on HeLa cells, see Figure 4.6), suggesting HCT116 cells in general are more sensitive to DNA damage than HeLa cells (*i.e.* the DNA replication inhibition that leads to histone mRNA degradation occurs faster in HCT116 cells).

In seeming contrast to above results, the substantial DNA damage-induced accumulation of H2A.X protein in wild-type HeLa cells (∼5-fold) is indiscernible in wild-type HCT116 cells (Figure 5.6A and compare wild-type cells with and without etoposide treatment in B (HeLa) and C (HCT116)). This difference is also reflected in the accumulation of γH2A.X, which is much higher in HeLa than HCT116 cells (Figure 5.6F, compare untreated and etoposide-treated samples in HeLa and HCT116 cells).

In both HeLa and HCT116 cells, knocking out the H2A.X SL mRNA (dHDE) leads to decreased total H2A.X and γH2A.X protein levels (Figure 5.6A and F, respectively). In HeLa cells, the decrease in total H2A.X protein is less pronounced (∼25% reduction in dHDE compared to wild-type, Figure 5.6D) than in HCT116 cells (∼70% reduction in dHDE compared to wild-type, Figure 5.6E), which correlates with the difference in H2A.X SL mRNA levels in the two wild-type cell lines (Figure 5.3). Accordingly, γH2A.X accumulation upon etoposide treatment is also less pronounced in HeLa dHDE compared to HCT116 dHDE cells (Figure 5.6F, compare G and H).

The H2A.X SL mRNA accumulation is S-phase-specific. Thus to distinguish the effect that the H2A.X SL mRNA "knock out" has on DNA damage signalling in S-phase and non-S-phase cells, γH2A.X signalling in HeLa (Figure 5.7A) and HCT116 (Figure 5.8C) wild-type and dHDE cells was analysed by immunofluorescence. Interestingly, unlike HeLa

cells, HCT116 cells show higher γH2A.X signal in S-phase cells compared to non-S-phase cells. This difference is reduced in dHDE compared to wild-type cells (Figures 5.7B and 5.8D, compare black and gray bars), likely due to the higher H2A.X SL mRNA expression in HCT116 cells in S-phase (Figure 5.4).

Strinkingly, in both cell lines, the absence of H2A.X SL mRNA leads to a decrease in γH2A.X levels in both S-phase and non-S-phase cells. This effect is milder in HeLa cells (no statistically significant difference in γH2A.X signal calculated from the immunofluorescence images, however significant in the western blot) but more pronounced in HCT116 cells (statistically significant in all experiments) (Figures 5.7B and 5.8D). Again, this correlates with levels of H2A.X SL mRNA in the corresponding wild-type cell line. The fact that non-S-phase dHDE cells that still contain the H2A.X poly(A)$^+$ mRNA also show decreased γH2A.X signal upon 2 hr etoposide treatment suggests that γH2A.X signalling does not solely depend on *de novo* H2A.X protein synthesis but rather also on the H2A.X protein already present in the chromatin (mainly incorporated during DNA replication in S-phase). Supporting this idea and in agreement with the low amount of H2A.X poly(A)$^+$ mRNA relative to H2A.X SL mRNA in HCT116 cells compared to HeLa cells, the depletion of the H2A.X poly(A)$^+$ mRNA in HCT116 cells did not lead to a significant decrease in either H2A.X or γH2A.X levels (Supplementary Figures S8 and S9). Overall, these results suggest that also the H2A.X protein incorporated into the chromatin during the S-phase of the cell cycle is important for DNA damage signalling.

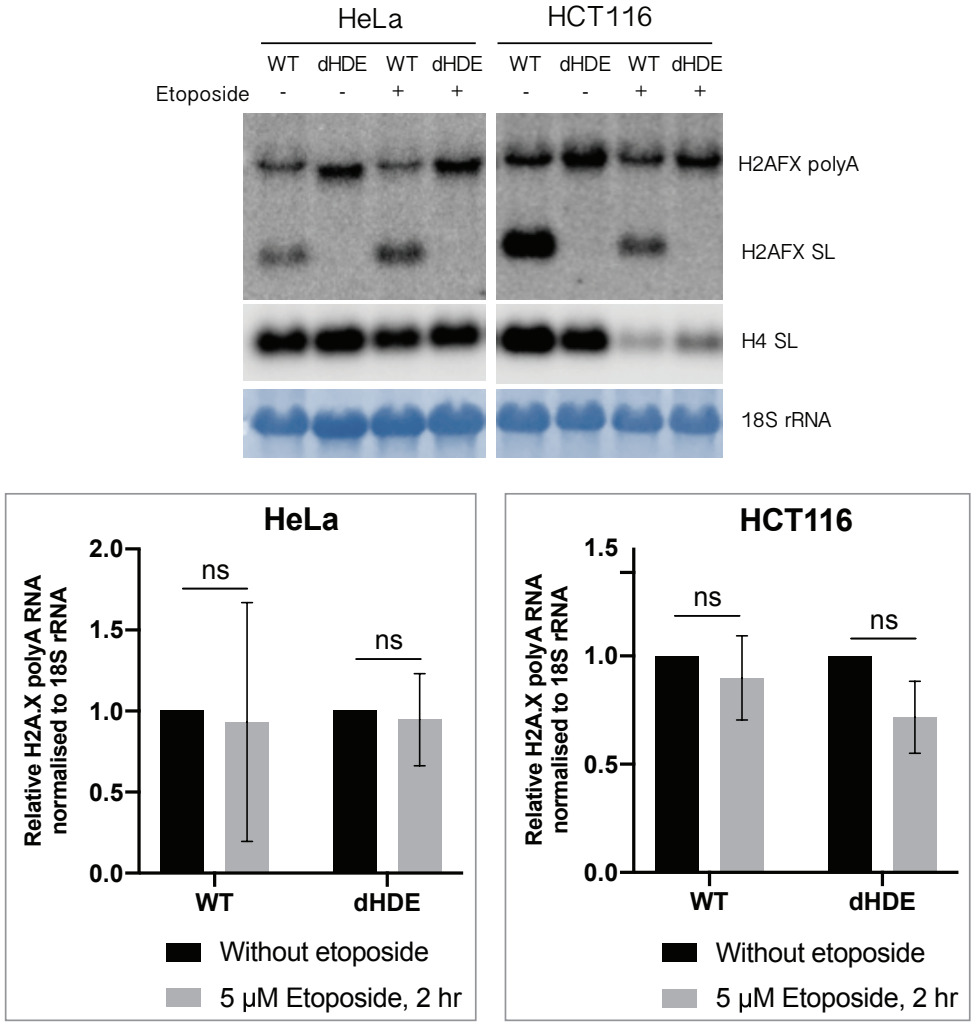



**Figure 5.5 Effect of DNA damage on H2A.X mRNA isoforms in H2A.X SL mRNA "knock out" cells.** Asynchronously growing wild-type (WT) and dHDE HeLa and HCT116 cells were treated with 5 µM etoposide for 2 hr before harvesting. Total RNA was used for northern blotting (this image shows a representative blot of three independent experiments). H2A.X SL mRNA and H2A.X poly(A)$^+$ mRNA isoforms were detected with a probe complementary to the common region between both mRNA isoforms. H4 SL: canonical H4 SL mRNA. 18S rRNA: 18S ribosomal RNA (methylene blue staining). The panels on the bottom show the quantification of the northern blot by densitometry of three independent experiments. Statistical significance was calculated by two-way Anova. ns: not significant, p-value $> 0.05$.

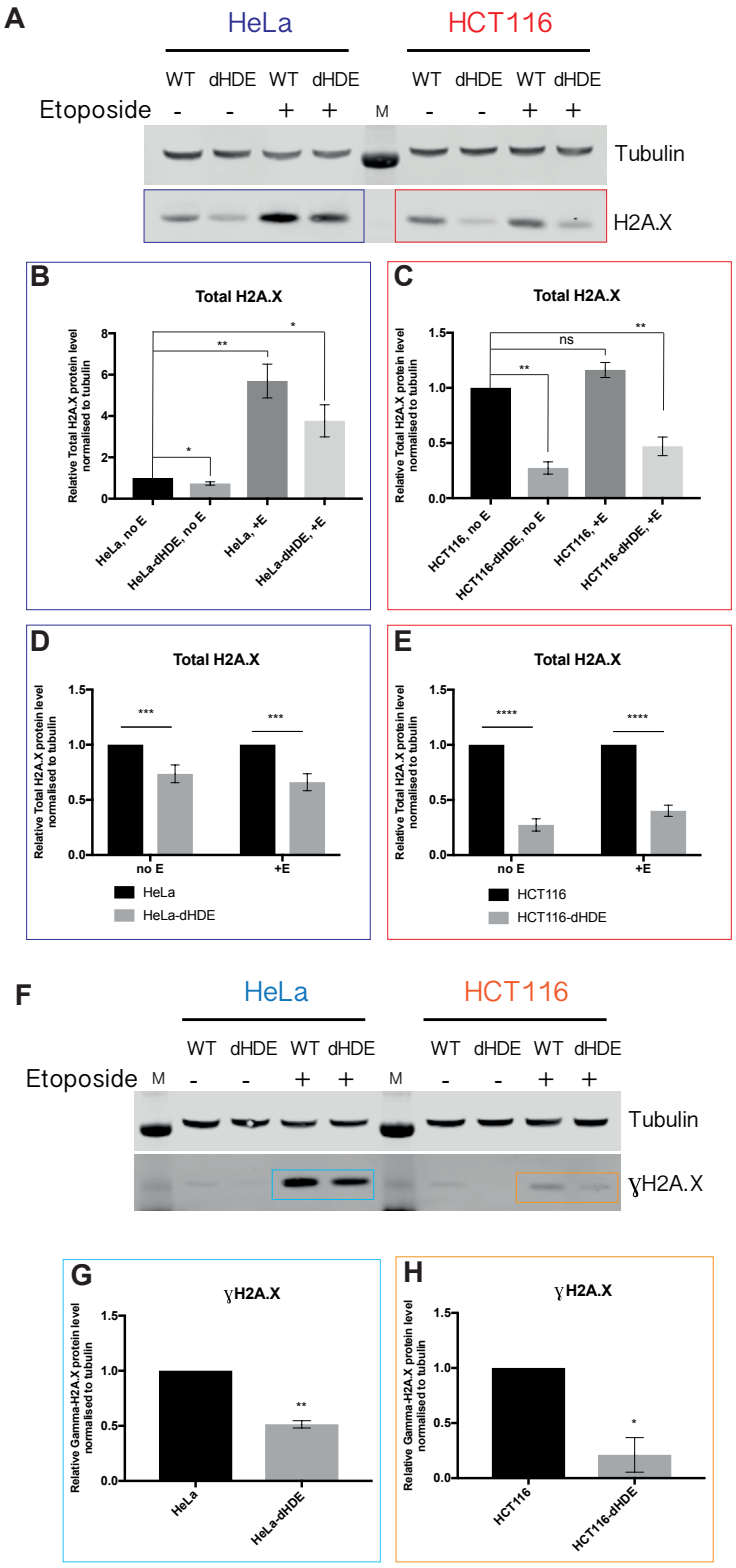



**Figure 5.6** (Legend on next page)

**Figure 5.6 H2A.X SL mRNA "knock out" leads to decreased H2A.X and γH2A.X levels in asynchronously growing cells.** HeLa and HCT116 wild-type and dHDE cells were treated with 5 µM etoposide for 2 hr before harvesting and total protein extracts were analysed by western blotting. A) Western blot showing total H2A.X protein levels. Tubulin was used as loading control. Total H2A.X protein levels shown in A were quantified by densitometry using the ImageStudioLite software: B) HeLa H2A.X levels normalised to tubulin and relative to untreated wild-type cells, C) HCT116 H2A.X levels normalised to tubulin and relative to untreated wild-type cells, D) HeLa H2A.X levels normalised to tubulin, in dHDE relative to wild-type cells, E) HCT116 H2A.X levels normalised to tubulin, in dHDE relative to wild-type cells. F) Western blot showing γH2A.X protein levels. Tubulin was used as loading control. γH2A.X protein levels shown in F were quantified by densitometry using the ImageStudioLite software: G) γH2A.X levels in HeLa cells treated with etoposide, normalised to tubulin and represented relative to the levels in untreated cells, H) γH2A.X levels in HCT116 cells treated with etoposide, normalised to tubulin and represented relative to the levels in untreated cells. The error bars represent the SE in three independent experiments, in B, C, G and H a paired t-test was used to calculate the statistical significance, whereas in D and E two-way Anova was used (Graphpad Prism software). ns: not significant P-value $> 0.05$, *P-value $\leq 0.05$, ** P-value $\leq 0.01$, *** P-value $\leq 0.001$, **** P-value $\leq 0.0001$. WT: wild-type cells. dHDE: CRISPR 9-nucleotide HDE deltetion cells. M: protein size marker. no E: no etoposide treatment. +E: 5 µM etoposide treatment for 2h before harvesting.

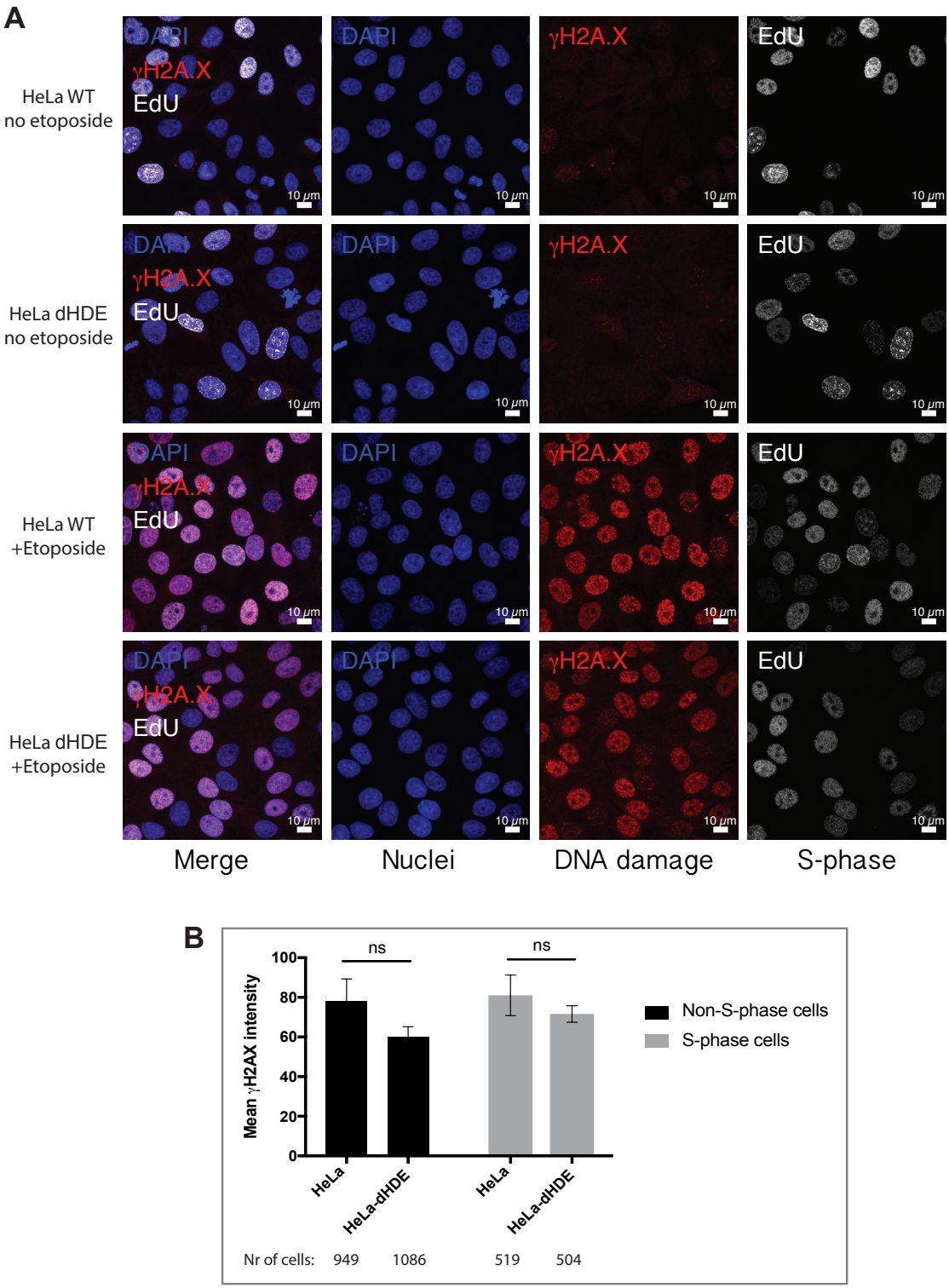



**Figure 5.7 H2A.X incorporation into the chromatin in S-phase is required for efficient γH2A.X signalling in HeLa cells.** Immunofluorescence images of HeLa (A). Wild-type and dHDE cells were treated with 5 µM etoposide for 2 hr, labelled with EdU (white) to distinguish the S-phase cells, stained with DAPI (blue) to visualise the nuclei and with γH2A.X antibody (red) as DNA damage indicator. B) Quantification of the nuclear γH2A.X signal in etoposide-treated HeLa cells, comparing the γH2A.X intensity between wild-type and dHDE cells in either non-S-phase or S-phase cells. 10 images of each sample were analysed for a total of three independent experiments. The total number (Nr) of cells analysed is indicated below the graph. The error bars represent the SE between the three independent experiments. Statistical significance was calculated by two-way Anova using the Graphpad Prism software. ns: not significant P-value > 0.05. The image analysis was performed using a macro in ImageJ to ensure unbiased quantification.

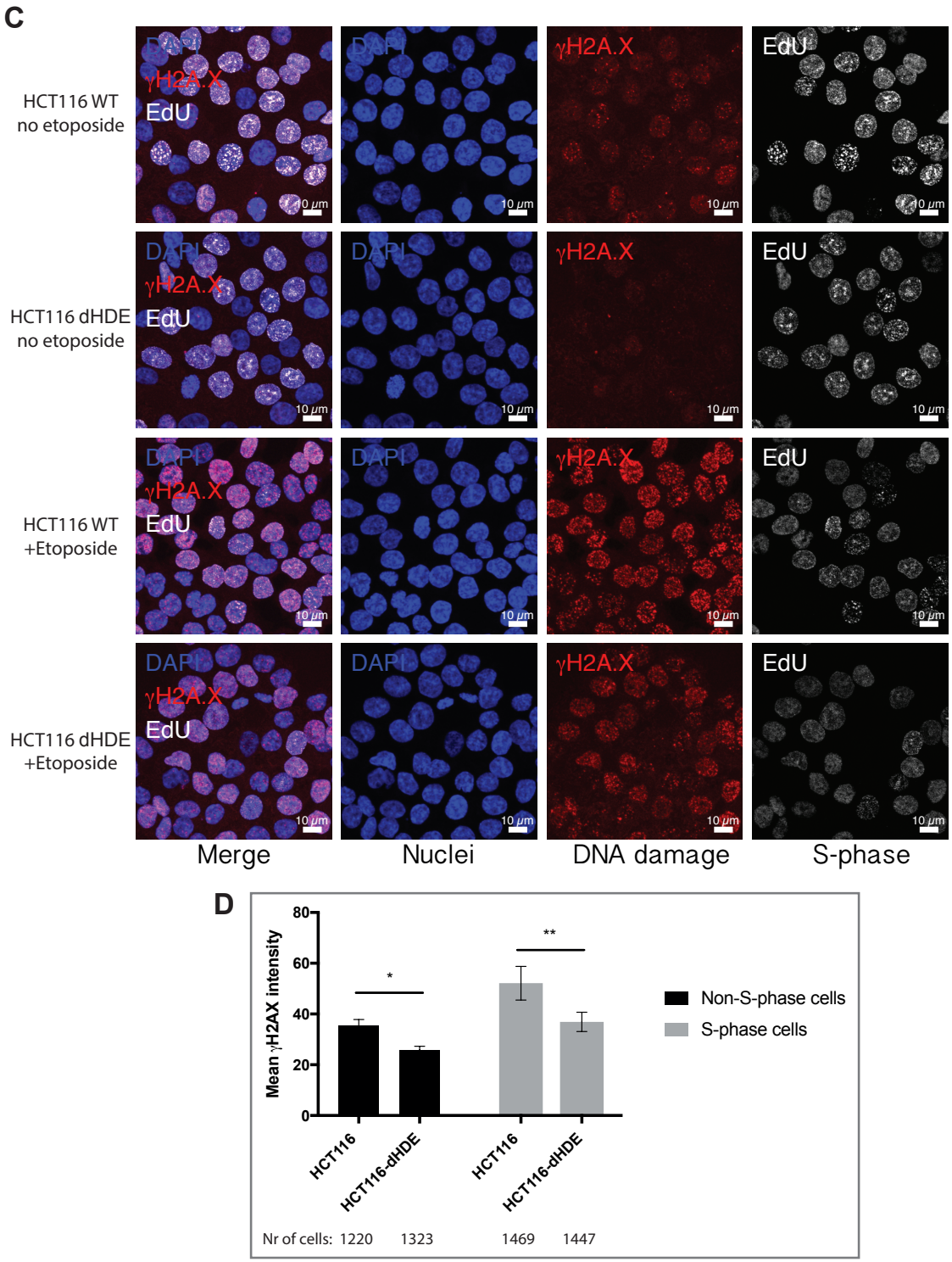



**Figure 5.8 H2A.X incorporation into the chromatin in S-phase is required for efficient γH2A.X signalling in HCT116 cells.** Immunofluorescence images of HCT116 (C) cells. The figure letters are a continuation of Figure 5.7. Wild-type and dHDE cells were treated with 5 µM etoposide for 2 hr, labelled with EdU (white) to distinguish the S-phase cells, stained with DAPI (blue) to visualise the nuclei and with γH2A.X antibody (red) as DNA damage indicator. D) Quantification of the nuclear γH2A.X signal in etoposide-treated HCT116 cells, comparing the γH2A.X intensity between wild-type and dHDE cells in either non-S-phase or S-phase cells. 10 images of each sample were analysed for a total of three independent experiments. The total number (Nr) of cells analysed is indicated below the graph. The error bars represent the SE between the three independent experiments. The error bars represent the SE between the three independent experiments. Statistical significance was calculated by two-way Anova using the Graphpad Prism software. ns: not significant P-value $> 0.05$, *P-value $\leq 0.05$, ** P-value $\leq 0.01$. The image analysis was performed using a macro in ImageJ to ensure unbiased quantification.

# Chapter 6

# Regulatory elements in the H2A.X poly(A)$^+$ mRNA

The SL structure of H2AFX not only ensures the existance of the H2A.X SL mRNA because it is essential for U7-dependent processing, but it also regulates multiple downstream steps in the metabolism of the H2A.X SL mRNA, mainly through SLBP. As the SL structure is also present in the H2A.X poly(A)$^+$ mRNA and it was shown that SLBP also binds the SL of the H2A.X poly(A)$^+$ mRNA [174], it is possible that SLBP has a regulatory function on this mRNA isoform. Furthermore, SLBP was suggested to indirectly bind to a highly conserved region in the H2A.X 3'-UTR close to the polyadenylation site (PAS) [174], further suggesting that SLBP may affect H2A.X poly(A)$^+$ mRNA metabolism.

To study the function of the SL structure in H2A.X poly(A)$^+$ mRNA, mutant H2AFX transgenes were constructed and integrated into HeLa Flp-In T-REx cells. The endogenous H2AFX was amplified by PCR from genomic DNA, including its own promoter and the entire extended 3'-UTR region past the PAS. This PCR product was then cloned into a plasmid containing the required FRT site for stable integration into the HeLa Flp-In T-REx cell line. The plasmid was modified further to create different H2A.X 3'-UTR mutants. Each of the H2AFX constructs was then integrated into the HeLa Flp-In T-REx cells. To distiguish the transgene from the endogenous H2AFX by PCR, a silent point mutation was included in the H2AFX ORF. See Chapter 2, Subsection 2.5.5, for the details of how these constructs were made.

Figure 6.1 shows a schematic of the different H2A.X pre-mRNA isoforms expressed from the integrated transgenes. #0 corresponds to the parental HeLa Flp-In T-REx cell line that does not contain any transgene and thus only expresses the endogenous H2A.X mRNA. #1 contains a wild-type H2AFX copy with the HindIII site in the ORF. #2 additionally contains the 9-nucleotide deletion of the HDE described in Chapter 5. In #3 the entire SL structure is deleted. #4 corresponds to the deletion of the conserved region in the extended 3'-UTR close to the PAS. #5 contains both SL and HDE deletions. #6 is deleted of both HDE and the conserved region and construct #7 contains a wild-type siRNA-resistant H2AFX gene (deletion of the "single siRNA" binding site).

For genotyping of the different cell lines, genomic DNA was amplified by PCR and the resulting PCR product digested with HindIII and sequenced. This experiment not only allowed genotype confirmation but also indicated that the HeLa Flp-In T-REx cells contain two endogenous H2AFX gene copies. Matthias Gromeier, who kindly provided the HeLa Flp-In T-REx cell line, confirmed the presence of a unique FRT site in these cells. Notably, the observed 2:1 endogenous:transgene ratio indicates the presence of two endogenous H2AFX genes in these cell lines (Supplementary Figure S10). Therefore, cell lines #1 to #7, contain two endogenous H2AFX copies and an additional transgene H2AFX copy.

To analyse the effect of these different mutations on pre-mRNA processing, total, poly(A)$^+$ and poly(A)$^-$ RNA was isolated from each of the 8 cell lines and analysed by RT-PCR with primers amplifying either the total H2A.X mRNA (SL and poly(A)$^+$) or specifically the H2A.X poly(A)$^+$ mRNA. The resulting PCR products were digested with HindIII to distinguish endogenous and transgene RNA (Figure 6.2). Initially, RNA was isolated from asynchronously growing cells. Uncut PCR products are shown to confirm that additional bands are specific products of HindIII digestion and not due to unspecific amplification (Figure 6.3, compare uncut and HindIII-digested PCR products). As expected due to the absence of a HindIII site in the endogenous H2AFX gene, the PCR product from parental HeLa Flp-In T-REx cells (no transgene) was resistant to HindIII digestion (Figure 6.3, compare uncut and HindIII digestion in cell line #0). All PCR products from cells that carry the HindIII site-containing transgene were effectively digested (tested but data not shown) with the enzyme (Figure 6.3, compare uncut and HindII digestion in cell lines #1 to #7). A control PCR was performed on "minus-RT" DNA (reverse transcription

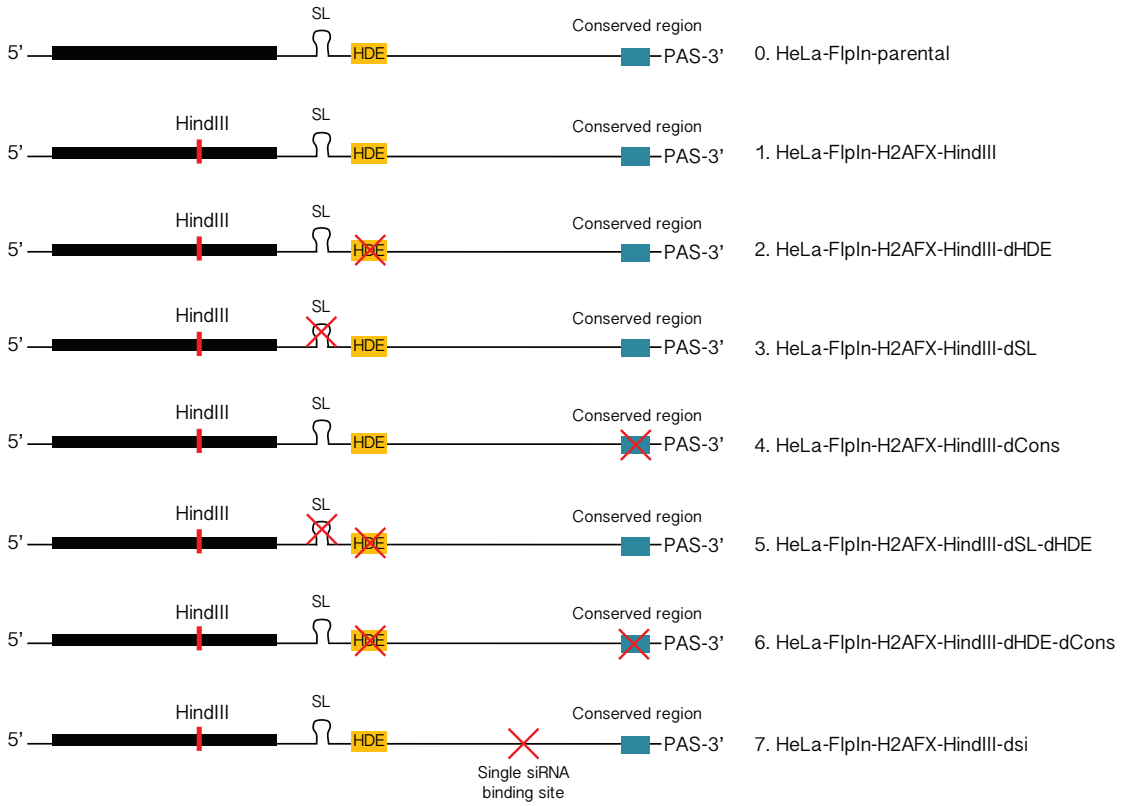



**Figure 6.1 Schematic showing the transgene H2A.X pre-mRNA expressed in the different HeLa Flp-In T-REx cell lines.** The cell lines are numbered from 0 to 7 (The red cross indicates deletion): #0: wild-type H2AFX gene. #1: wild-type H2AFX gene with a silent point mutation in the ORF creating a HindIII site. #2: HindIII site and 9-nucleotide deletion of the HDE. #3: HindIII site and SL deletion, #4: HindIII site and deletion of the conserved region. #5: HindIII site, SL and HDE deletion. #6: HindIII site, SL and conserved region deletion. #7: HindIII site and deletion of the "single siRNA" binding site (siRNA resistant).

reaction without reverse transcriptase) to confirm the absence of genomic DNA in the RNA samples used for reverse transcription (Figure 6.3, "Minus RT").

Primers binding to the common region between both H2A.X mRNA isoforms (Total H2A.X primers) amplify both mRNA isoforms if cDNA from total RNA is used for the PCR. However, if cDNA from poly(A)$^-$ or poly(A)$^+$ RNA (hereinafter referred to as poly(A)$^-$ and poly(A)$^+$ cDNA) is used, Total H2A.X primers mainly amplify the non-polyadenylated H2A.X SL mRNA and the H2A.X poly(A)$^+$ mRNA, respectively. The primers that bind to the extended 3'-UTR (H2A.X poly(A)$^+$ primers) specifically amplify the H2A.X poly(A)$^+$ mRNA. These will only generate a PCR product from poly(A)$^-$ cDNA if there is remaining "contaminating" poly(A)$^+$ RNA in the poly(A)$^-$ fraction, thus indicative of inefficient fractionation (*e.g.* see PCR product generated with H2A.X poly(A)$^+$ primers

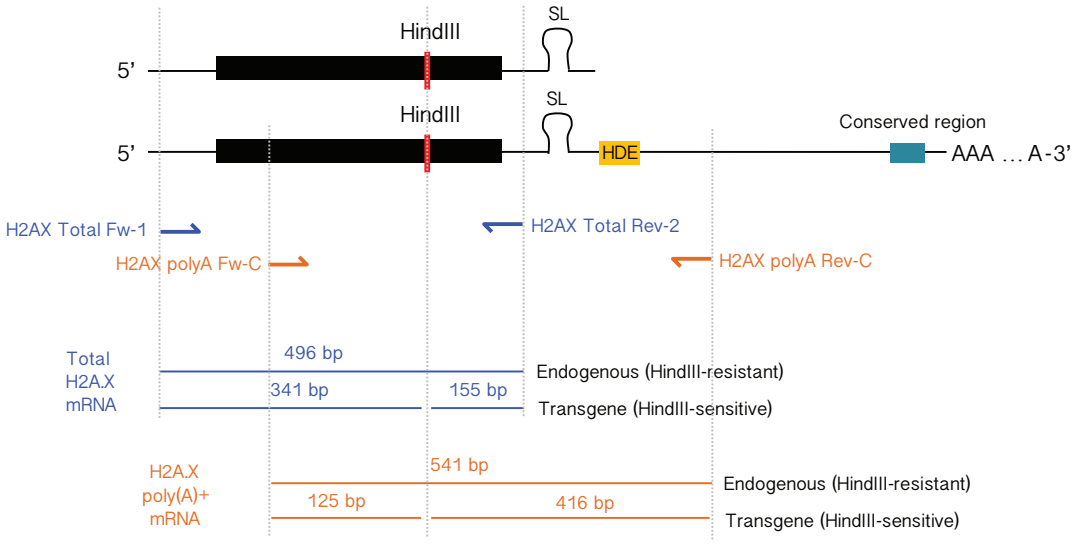



**Figure 6.2 PCR approach to distinguish transgene and endogenous H2AFX.**

from poly(A)$^-$ cDNA from cell line #0 in Figure 6.3, orange arrows).

From the experiment performed on asynchronously growing cells, it is immediately apparent that the HDE is required for efficient H2A.X pre-mRNA processing at the U7-dependent site: reduced transgene H2A.X SL mRNA was detected whenever the HDE is deleted (Figure 6.3, compare cell lines #2, #5 and #6 (blue circles) with #1, #3, #4 and #7 in the poly(A)$^-$ cDNA RT-PCR performed with Total H2A.X primers). Surprisingly, the SL deletion still showed some transgene H2A.X mRNA in the poly(A)$^-$ sample (Figure 6.3, cell line #3, poly(A)$^-$, pink circle), suggesting the SL is inessential for processing at the U7-dependent site. But this hypothesis was contradicted by a similar PCR experiment where size-selected RNA was used instead of poly(A)-selected RNA (Figure 6.4), confirming that whenever the SL or the HDE is deleted, U7-dependent processing is prevented and no H2A.X SL mRNA is generated. Therefore, the poly(A)$^+$ selection was inefficient in some samples.

Interestingly, the deletion of the conserved region close to the PAS leads to decreased accumulation of transgene H2A.X poly(A)$^+$ mRNA relative to the endogenous (Figure 6.3, see cell line #4 in total and poly(A)$^+$ RT-PCRs using primers specifically amplifying H2A.X poly(A)$^+$ mRNA (right part of the gel), light green arrows). This suggests a role for the conserved region in polyadenylation or mRNA stability.

The dHDE CRISPR cells allowed assessment of the effect of the absence of the H2A.X SL mRNA on the cell cycle distribution of H2A.X poly(A)$^+$ mRNA (Figure 5.4). This experiment showed that the H2A.X poly(A)$^+$ mRNA is only mildly upregulated in S-phase when U7-dependent processing is abolished, suggesting that there is a control mechanism that regulates H2A.X poly(A)$^+$ mRNA levels. As the H2A.X poly(A)$^+$ mRNA produced in dHDE cells contains the SL, this structure may be involved in controlling H2A.X poly(A)$^+$ mRNA stability in S-phase. To test whether the deletion of the SL changes the H2A.X poly(A)$^+$ mRNA cell cycle distribution in a different manner to the HDE deletion, a DTB was performed on the HeLa Flp-In T-REx cell lines #1, #2, #3 and #4 (H2AFX-HindIII, H2AFX-HindIII-dHDE, H2AFX-HindIII-dSL and H2AFX-HindIII-dCons, see Supplementary Figure S12 for flow cytometry analysis)) and the production of H2A.X SL mRNA and H2A.X poly(A)$^+$ mRNA was assessed by PCR. Initially the RT-PCR was performed on total RNA, with either Total H2A.X primers or H2A.X poly(A)$^+$ primers (Figure 6.5). If the cell cycle regulation of the endogenous and the transgene H2AFX genes was the same, the transgene:endogenous ratio would be constant during the cell cycle. However, this is not the case (Figure 6.5, see cell line #1 (H2AFX-HindIII), red arrow), likely due to the different genomic location of the transgene. Even though the transgene H2AFX is under the control of the endogenous H2AFX promoter, it is possible that other regions around the endogenous H2AFX locus influence endogenous H2AFX expression that do not affect the transgene. Nevertheless, the expression of the different transgene H2AFX genes is regulated in the same way, which allows comparison between the different transgene mutants. Importantly, the PCR cycle number has no influence on the transgene:endogenous ratio (Supplementary Figure S13). This suggests that the ratio derived from the RT-PCR data is quantitative. Furthermore, the "minus-RT" of DTB-synchronised samples showed the DNAse digestion was efficient, confirming the amplification specifically results from cDNA and not genomic DNA (Supplementary Figure S11).

Deletion of both the HDE and the SL lead to less total H2A.X mRNA in S-phase, as the transgene:endogenous ratio at the 0 to 5 hr time points is decreased compared to the wild-type (Figure 6.5, compare samples highlighted in green boxes in cell lines #2 and #3 with #1 in the upper panel (Total H2A.X primers)). This is expected since these deletions abolish U7-dependent processing and thus transgene H2A.X SL mRNA production. The

transgene:endogenous ratio in the remaining time points is similar between the four cell lines and it is difficult to draw conclusions about the SL and poly(A)$^+$ H2A.X mRNA levels without fractionating the DNA or using specific primers. Thus to specifically determine transgene H2A.X poly(A)$^+$ mRNA expression levels, the PCR was performed using H2A.X poly(A)$^+$ primers (Figure 6.5, lower panel). This PCR shows that both HDE and SL deletions increase H2A.X poly(A)$^+$ mRNA levels to some extent when cells enter the S-phase as the transgene:endogenous ratio is increased compared to the wild-type (Figure 6.5, lower panel, compare cell lines #2 and #3 with #1, blue boxes). The same effect is observed when the PCR is performed on poly(A)$^+$ cDNA using either Total or poly(A)$^+$ H2A.X primers (Figure 6.6 and 6.7, blue boxes).

Given the fact that the HDE and the SL are essential for H2AFX U7-dependent processing (Figure 6.4), the transgene bands observed in dHDE and dSL samples in the poly(A)$^-$ cDNA PCR performed with Total H2A.X primers represent the inefficiency of the poly(A)-selection (Figure 6.6, see dHDE and dSL poly(A)$^-$). Similarly, the amplification of poly(A)$^-$ cDNA using H2A.X poly(A)$^+$ primers represent the efficency of the poly(A)$^+$ RNA depletion from poly(A)$^-$ RNA (Figure 6.7, poly(A)$^-$ RNA).

Figure 6.8 shows enlarged HindIII-digested PCR products from Figure 6.7 to simplify the comparison of transgene H2A.X poly(A)$^+$ mRNA cell cycle regulation in the different mutants. The H2A.X poly(A)$^+$ mRNA levels generated from the transgene dHDE and dSL mutants specifically decrease at the 5 hr time point, more so for dHDE than dSL (see red box in Figure 6.8). At 5 hr ~90% of the cells are in S-phase (Supplementary Figure S12), suggesting a role of the SL in restricting H2A.X poly(A)$^+$ mRNA accumulation in S-phase. This effect in the 5 hr time point is also observed in Figures 6.5 (lower panel, compare 5 hr time points in dHDE and dSL cells, red boxes) and 6.6 (compare 5 hr time point in dHDE and dSL, red boxes).

As shown in the experiment using asynchronously growing cells (Figure 6.3), the deletion of the conserved region consistently leads to decreased accumulation of H2A.X poly(A)$^+$ mRNA throughout the cell cycle, but especially in S-phase and G2/M (cell line #4 in Figures 6.5 (lower panel), 6.6, 6.7 and 6.8).

Taken together, consistent with data obtained from the synchronised dHDE CRISPR cells, the deletion of both HDE and SL somewhat increased H2A.X poly(A)$^+$ mRNA levels

in S-phase. Furthermore, the SL may be involved in regulating the stability of the H2A.X poly(A)$^+$ mRNA in S-phase. Finally, the conserved region regulates H2A.X poly(A)$^+$ mRNA levels, either by influencing polyadenylation or H2A.X poly(A)$^+$ mRNA stability.

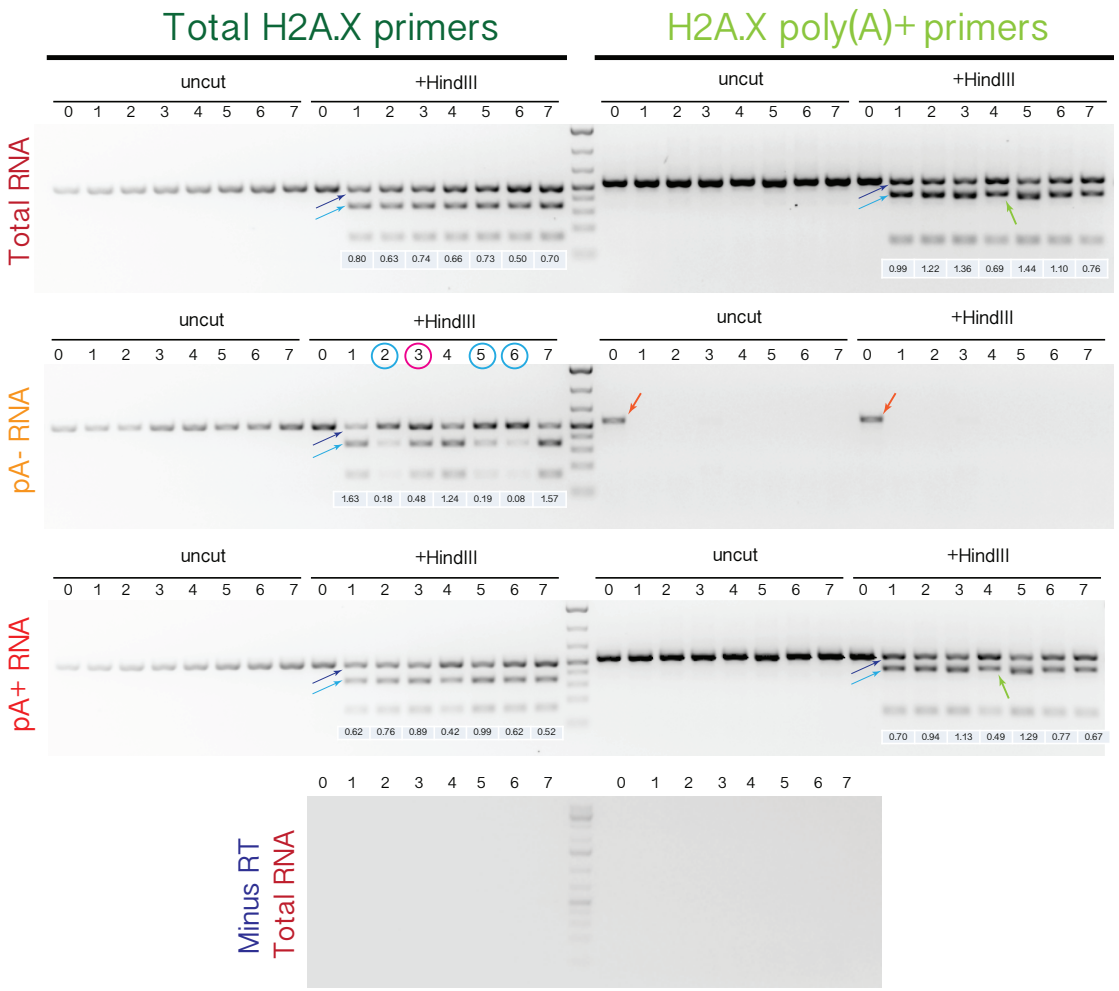



**Figure 6.3 Both SL and HDE are essential for processing at the H2AFX U7-dependent site and the conserved region close to the PAS is required for efficient acumulation of H2A.X poly(A)$^+$ mRNA.** Total RNA was extracted from the 8 different HeLa Flp-In T-REx cell lines (Figure 6.1) and subjected to poly(A)$^+$ selection. Total, poly(A)$^+$ and poly(A)$^-$ RNA was used for RT PCR with primers amplifying either the total (SL and poly(A)$^+$) or specifically the H2A.X poly(A)$^+$ mRNA. PCR products were digested with HindIII. Uncut and HindIII-digested PCR products were loaded on an ethidium bromide-containing 2% agarose gel. The numbers below each lane indicate the transgene:endogenous ratio. Representative picture of three independent experiments. Dark blue arrow: endogenous. Light blue arrow: transgene. The orange arrows mark bands resulting from remaining "contaminating" poly(A)$^+$ RNA in the poly(A)$^-$ fraction, thus indicative of inefficient fractionation in this particular sample. Light green arrows point to the band resulting from decreased transgene H2A.X poly(A)$^+$ mRNA in the cell line where the conserved region close to the PAS was deleted. Complete HindIII digestion was confirmed (data not shown).

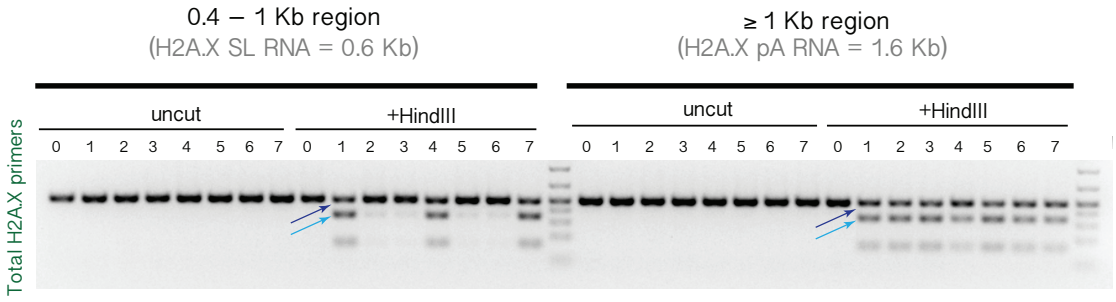



**Figure 6.4 The HDE and the SL are essential for H2AFX U7-dependent processing.** Total RNA from asynchronously growing HeLa Flp-In T-REx cell lines (Figure 6.1) was separated on a 6% TBE-Urea gel. The regions of the gel comprising RNA sizes from 0.4 to 1 Kb (containing the 0.6 Kb H2A.X SL mRNA) and larger than 1 Kb (containing the 1.6 Kb H2A.X poly(A)$^+$ mRNA) were excised, and the eluted RNA was used for RT-PCR with primers amplifying total H2A.X mRNA (SL and poly(A)$^+$). PCR products were digested with HindIII. Uncut and HindIII-digested PCR products were loaded on an ethidium bromide-containing 2% agarose gel. Dark blue arrow: endogenous. Light blue arrow: transgene.

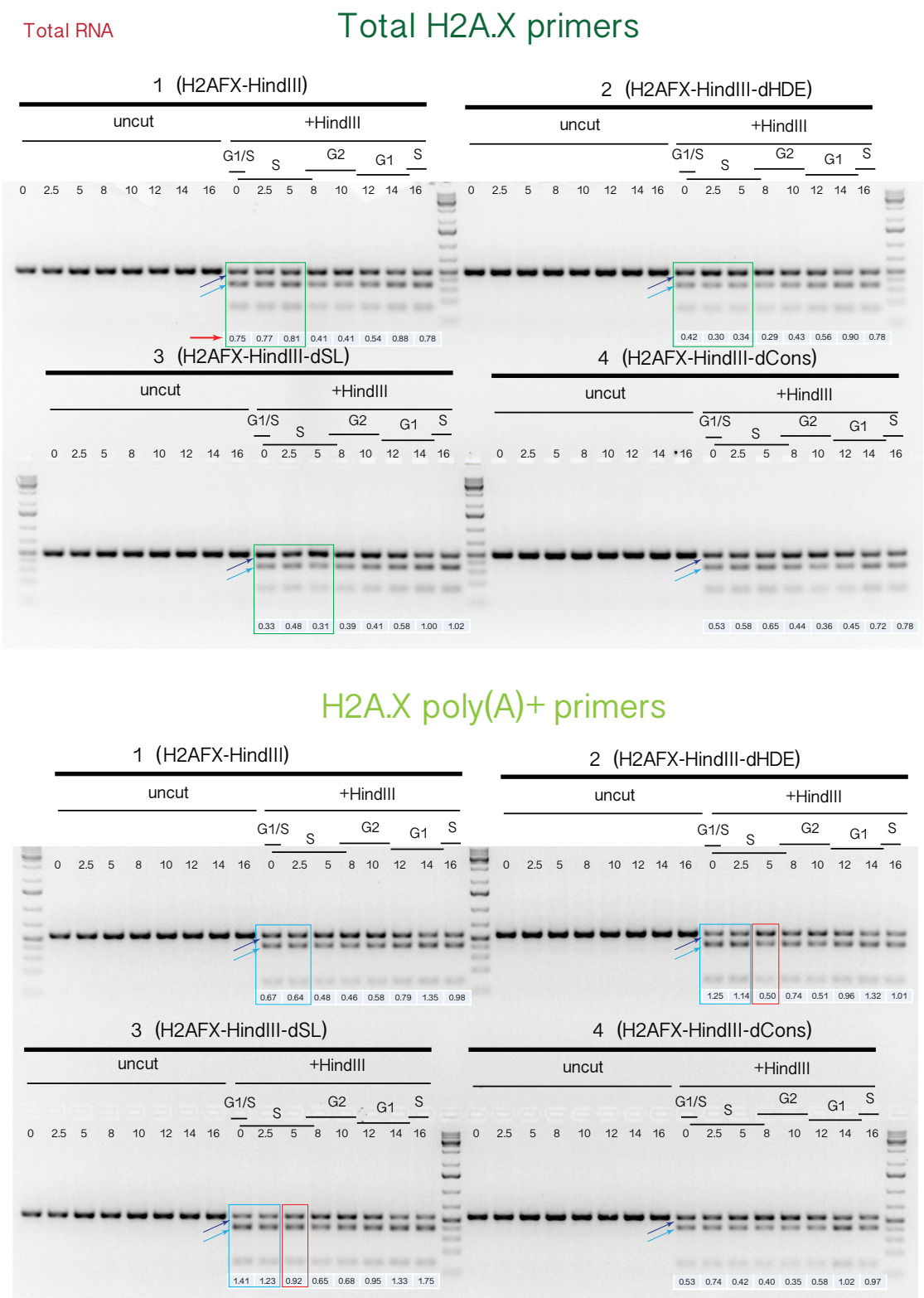



**Figure 6.5 Total RNA RT-PCR of DTB-synchronised HeLa Flp-In T-REx cell lines.** HeLa Flp-In T-REx cell lines #1, 2, 3 and 4 (H2AFX-HindIII, H2AFX-HindIII-dHDE, H2AFX-HindIII-dSL and H2AFX-HindIII-dCons) were synchronised by DTB and total RNA was used for RT-PCR. The images represent the uncut and HindIII-digested PCR products generated with Total H2A.X primers (top panel) and H2A.X polyA primers (lower panel) loaded on an ethidium bromide-stained 2% agarose gel. The numbers on top of each lane indicate the time in hr after the release from the DTB. The letters indicate the corresponding cell cycle phase determined by flow cytometry (Supplementary Figure S12). The numbers in the box below each lane indicate the transgene:endogenous ratio calculated by densitometry using the ImageStudioLite software. Dark blue arrow: endogenous. Light blue arrow: transgene.

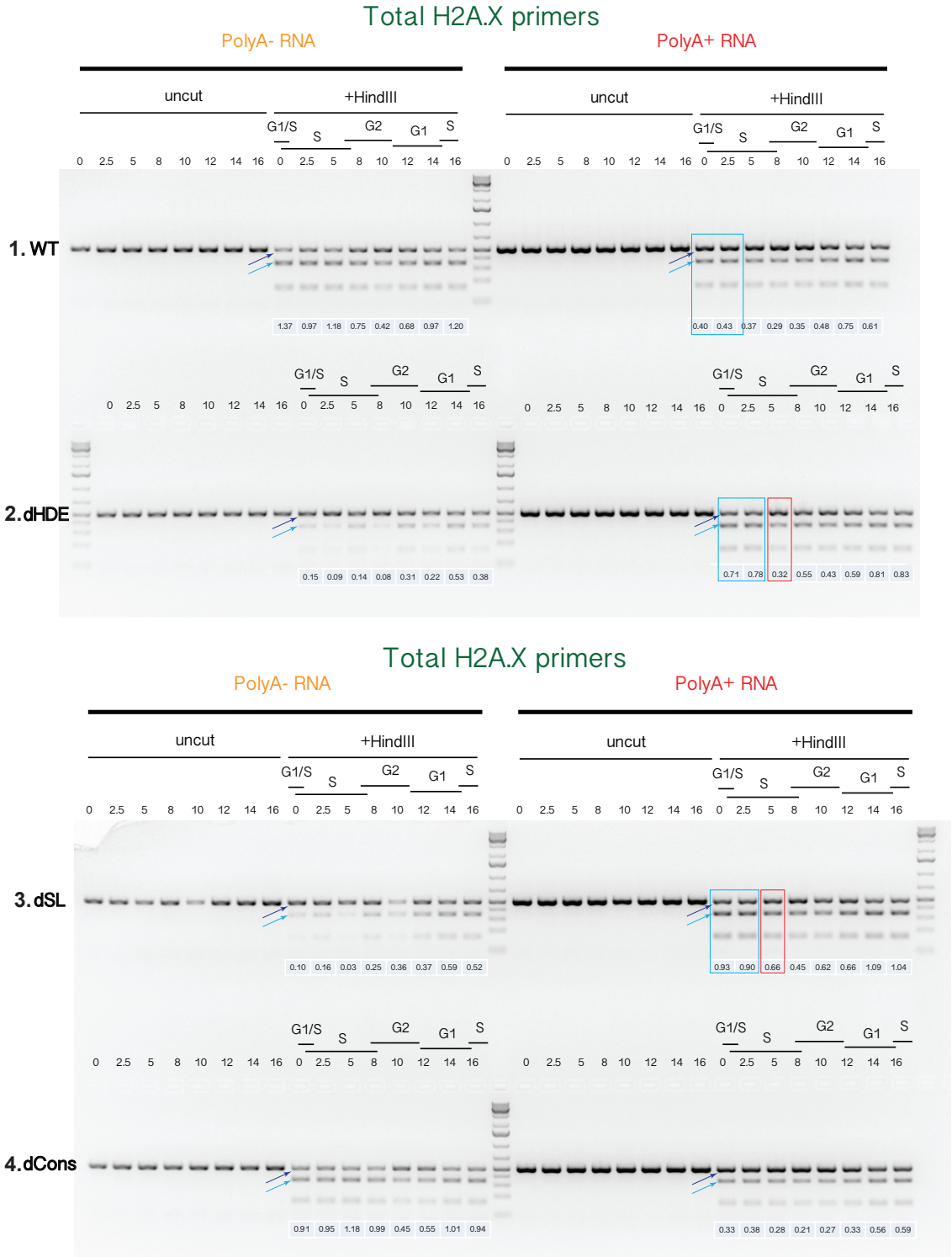



**Figure 6.6 Total H2A.X RT-PCR of DTB-synchronised HeLa Flp-In T-REx cell lines (poly(A)$^-$ and poly(A)$^+$ RNA).** HeLa Flp-In T-REx cell lines #1, 2, 3 and 4 (H2AFX-HindIII, H2AFX-HindIII-dHDE, H2AFX-HindIII-dSL and H2AFX-HindIII-dCons) were synchronised by DTB. The RNA was subjected to poly(A)-selection and poly(A)$^-$ and poly(A)$^+$ RNA was used for RT-PCR. The images represent the uncut and HindIII-digested PCR products generated with Total H2A.X primers loaded on an ethidium bromide-stained 2% agarose gel. The numbers on top of each lane indicate the time in hr after the release from the DTB. The letters indicate the corresponding cell cycle phase determined by flow cytometry (Supplementary Figure S12). The numbers in the box below each lane indicate the transgene:endogenous ratio calculated by densitometry using the ImageStudioLite software. Dark blue arrow: endogenous. Light blue arrow: transgene.

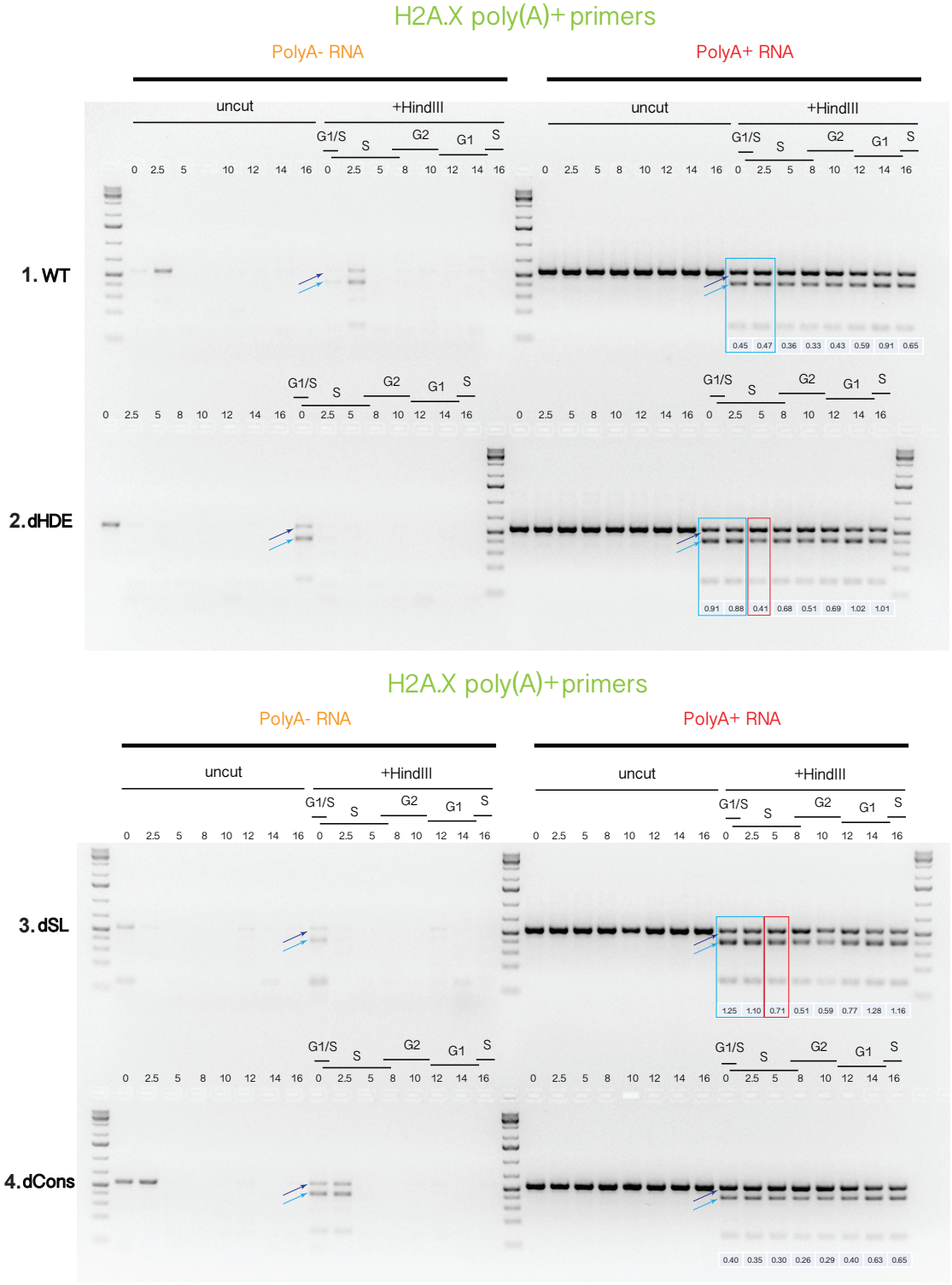



**Figure 6.7 Poly(A)$^+$ H2A.X RT-PCR of DTB-synchronised HeLa Flp-In T-REx cell lines (poly(A)$^-$ and poly(A)$^+$ RNA).** HeLa Flp-In T-REx cell lines #1, 2, 3 and 4 (H2AFX-HindIII, H2AFX-HindIII-dHDE, H2AFX-HindIII-dSL and H2AFX-HindIII-dCons) were synchronised by DTB. The RNA was subjected to selection and poly(A)$^-$ and poly(A)$^+$ RNA was used for RT-PCR. The images represent the uncut and HindIII-digested PCR products generated with poly(A)$^+$ H2A.X primers loaded on an ethidium bromide-stained 2% agarose gel. The numbers on top of each lane indicate the time in hr after the release from the DTB. The letters indicate the corresponding cell cycle phase determined by flow cytometry (Supplementary Figure S12). The numbers in the box below each lane indicate the transgene:endogenous ratio calculated by densitometry using the ImageStudioLite software. Dark blue arrow: endogenous. Light blue arrow: transgene.

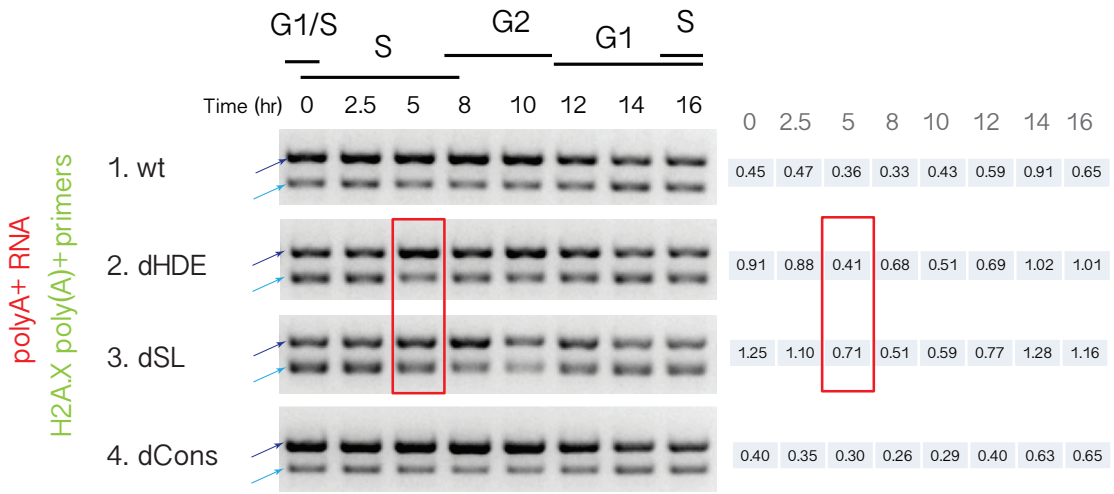



**Figure 6.8 Summary image from Figure 6.7.** Magnification of the poly(A)$^+$ RNA RT-PCR images shown in Figure 6.7. The numbers on top of each lane indicate the time in hr after the release from the DTB. The letters indicate the corresponding cell cycle phase determined by flow cytometry (Supplementary Figure S12). The numbers on the right of each image correspond to the transgene:endogenous ratio of each lane. Dark blue arrow: endogenous. Light blue arrow: transgene. This experiment was performed only once, repetitions are required to calculate statistics.

# Chapter 7

# Discussion

## 7.1 HeLa cells express polyadenylated RD histone mRNA

Differentiated cells that have a longer half-life than canonical histone proteins depend on so-called "replacement histones" to compensate for histone loss. Furthermore, it has been suggested that in somatic cells canonical histones displaced from actively transcribed loci can be degraded and replaced by *de novo* histone proteins instead of depending on recycling of old histones [325]. Differentiated cells tend to favour expression of RI histone variants, with the exception of H2b and H4 [251]. However, these variants have a different protein structure compared to their canonical histone counterpart and are therefore associated with specialised functions rather than serving as canonical replacement histones. It has previously been shown that replacement histones may be translated from polyadenylated RD histone mRNA, generated as part of the differentiation program [251]. However, the present study suggests that the production of poly(A)$^+$ RD histone mRNA is not only linked to differentiation, as cycling HeLa cell RD histone genes also express such mRNA isoforms (Table 3.1). About ∼50% of RD histone genes in HeLa cells express some level of polyadenylated mRNA. Interestingly, the genes in category 4 and 5 overlap with those found in differentiated cells, specifically HIST1H1C, HIST2H2AA3/4, HIST2H3BE (category 4), and the genes giving rise to spliced polyadenylated mRNA HIST1H2AC, HIST1H2BC, HIST1H2BD, HIST1H2BK, HIST1H4H and HIST2H4A/B (category 5) [251].

In HeLa cells, the poly(A)$^+$ mRNA isoforms expressed from RD histone genes are also detected in S-phase-enriched cells (Figure 3.4). As observed for H2AFX, this can

be a consequence of inefficient U7-dependent processing in S-phase allowing Pol II to reach the polyadenylation site (PAS) before processing at the U7-dependent site occurs. However as U7-dependent processing is highly efficient in S-phase, it is more likely that RD histone poly(A)$^+$ mRNA is mainly generated by the basal transcription occuring outside of S-phase when U7-dependent processing is inefficient [209]. The extended 3'-UTR and the poly(A)-tail may confer increased mRNA stability throughout the cell cycle, explaining their detection in S-phase.

The poly(A)$^+$ mRNA expressed from category 4 and 5 genes are also detected in the cytoplasm (Figure 3.4), suggesting they may be translated. A study showing poly(A)$^+$ histone mRNA in polyribosome fractions supports this hypothesis [326]. As these mRNA isoforms are present throughout the cell cycle, they may allow canonical histone protein production outside of S-phase, where they might serve as replacement of damaged histones. However, further study is required to establish such a function, especially in actively dividing cells.

## 7.2 Cell cycle regulation of the H2A.X mRNA isoforms

Even though the RNAseq analysis shown in Chapter 3, Section 3.1, is not fully quantitative, it is apparent that H2AFX is the histone gene that produces the highest levels of poly(A)$^+$ mRNA compared to its S-phase-specific counterpart (Figure 3.5). Since the late 1980s it has been known that the H2A.X poly(A)$^+$ mRNA is replication-independent as it is stable when DNA replication is inhibited [256]. However, its detailed cell cycle distribution had not been studied. As shown in Figures 3.6 and 3.7, in both HeLa and HCT116 cells the H2A.X poly(A)$^+$ mRNA is present throughout the cell cycle, peaking in G2/M. A high-throughput cell-cycle transcriptome analysis performed on HeLa cells supports this finding showing that H2AFX is in the list of genes upregulated during G2/M [327]. Higher H2A.X poly(A)$^+$ mRNA levels in G2/M might reflect an increased requirement for H2A.X protein at this stage of the cell cycle, *e.g.* it was shown that normally proliferating cells show DNA damage-independent fluctuations in $\gamma$H2A.X signal during the cell cycle, reaching maximum levels at M-phase [302]. However, the Esashi lab (University of Oxford) does not observe such $\gamma$H2A.X fluctuations during the cell cycle.

The H2A.X SL mRNA was shown to be only weakly linked to DNA replication, as its decay and synthesis rates are only partially affected by the inhibition of DNA and protein synthesis, respectively [256]. In contrast, my data demonstrates that H2A.X SL mRNA production is strongly upregulated in S-phase and downregulated as DNA replication concludes, in HeLa but especially in HCT116 cells (Figures 3.6, 3.7 and 5.4). The effect is less pronounced in RPE-1 cells even though these cells do express some S-phase-specific H2A.X SL mRNA (Figures 3.8 and 4.1F and G). The variation in SL *versus* poly(A)$^+$ mRNA levels in different cell lines had already been noticed [256], however it is unclear what the cause of this difference is. HeLa cells have only about 10% as much U7 and FLASH compared to other tumor cells, and normal cells (including RPE-1 cells) likely also have less (oral communication by William Marzluff). Therefore, it is possible that HeLa and RPE-1 cells do not recruit the histone 3'-end processing machinery fast enough to efficiently process at the U7-dependent site before Pol II reaches the PAS (inefficient U7-dependent processing). However, further study is required to confirm this hypothesis.

Unlike RD histone genes that are located in clusters on chromosomes 6 and 1, H2AFX is separately located on chromosome 11. This clustering of histone genes allows the highly coordinated expression of RD histone genes, through the formation of the histone locus body (HLB) that concentrates essential factors for the expression of RD histone genes around the histone gene locus [184]. The double thymidine block (DTB) synchronisation experiments performed on HeLa and HCT116 cells, show that H2A.X SL mRNA is already detected at the 0 hr time point, when SLBP levels are already increased. In contrast, RD histone mRNA such as H4 SL mRNA is not yet expressed Figures 3.6B and 3.7B). This suggests that the formation of H2A.X SL mRNA is regulated differently than RD histone genes located in clusters. H2A.X SL mRNA expresssion seems to mainly depend on the presence of histone 3'-end processing factors whereas RD histone mRNA expression is controlled by additional factors recruited to the HLB.

The fact that RPE-1 cells arrested in G0 and released to synchronously enter the cell cycle accumulate H2A.X poly(A)$^+$ mRNA in G1-phase not only proves that the H2AFX gene is transcribed outside of S-phase but also that this mRNA isoform is indeed generated outside of S-phase (Figure 3.8). Interestingly, these cells continue accumulating substantial amounts of H2A.X poly(A)$^+$ mRNA as they progress into S-phase (Figure 3.8). This

effect may be a consequence of recovery from G0 as at 22 hr after re-entering the cell cycle, nearly 50% of the cells are in S-phase whereas the other half remain in G1 and continues to accumulate H2A.X poly(A)$^+$ mRNA. Thus the H2A.X poly(A)$^+$ mRNA levels observed at 22 hr result from a combination of transcription in G1 and potentially inefficient U7-dependent processing in S-phase. The latter view is supported by the fact that HeLa and especially HCT116 cells also show some upregulation of H2A.X poly(A)$^+$ mRNA levels as cells progress through S-phase and particularly at the end of S-phase (Figure 5.4, see H2A.X poly(A)$^+$ mRNA levels in HeLa and HCT116 wild-type cells).

In conclusion, the H2AFX gene is transcribed throughout the cell cycle, as previously suggested [256]. The H2A.X SL mRNA is specifically generated in S-phase by U7-dependent processing, whereas the H2A.X poly(A)$^+$ mRNA is generated by CPA outside of S-phase and low levels may be produced in S-phase as a result of inefficient U7-dependent processing. Furthermore, the H2A.X poly(A)$^+$ mRNA is clearly upregulated in G2/M in both HeLa and HCT116 cells. Interestingly, HCT116 cells accumulate much higher levels of H2A.X SL *versus* poly(A)$^+$ mRNA compared to HeLa and RPE-1 cells, possibly due to different levels of 3'-end processing machinery components.

## 7.3 The H2A.X poly(A)$^+$ mRNA is required for efficient $\gamma$H2A.X signalling

The formation of DNA double-stranded breaks (DSBs) leads to the phosphorylation of H2A.X at Serine 139, forming $\gamma$H2A.X [259]. This modification has been shown to stimulate DNA damage repair by enabling stable association of DNA repair factors to break sites [272, 273]. The early finding that the H2A.X SL mRNA is only weakly linked to DNA replication and the fact that the H2A.X poly(A)$^+$ mRNA is expressed throughout the cell cycle [256], led to inclusion of the H2A.X protein into the group of replication-independent (RI) histone variants. As such, the H2A.X protein has been the focus of intensive study, especially in relation to its function in DNA damage repair. Almost all studies involving H2A.X depletion by RNAi target both H2A.X mRNA isoforms. To my knowledge, only three studies involve specific depletion of the H2A.X poly(A)$^+$ mRNA, focusing specifically on H2A.X regulation by miRNAs [311–313]. These publications show that the miRNA-dependent depletion of H2A.X poly(A)$^+$ mRNA can lead to decreased

H2A.X protein levels and genome instability. However, the regulation of H2A.X levels by miRNAs appears to be highly cell-type specific [312]. This variation between cell lines may be explained by differences in H2A.X mRNA isoform ratios in the different cell lines, as in the present study I show that cells expressing high levels of the H2A.X SL mRNA do not require the H2A.X poly(A)$^+$ mRNA for efficient $\gamma$H2A.X signalling and would therefore not be susceptible to miRNAs targeting the extended 3'-UTR. Furthermore, the cell cycle-specific $\gamma$H2A.X signalling was not determined in these studies.

Similar to the above mentioned miRNA studies [311–313], I show that depletion of the H2A.X poly(A)$^+$ mRNA by RNAi in HeLa and RPE-1 cells leads to decreased protein levels and to a defect in cell cycle progression, possibly due to the accumulation of unrepaired DNA damage (Figure 4.1). A recent publication shows that, in normal conditions, the H2A.X protein is constantly degraded through poly-ubiquitination in resting cells [322]. Upon DNA damage induction, this degradation pathway is blocked, allowing H2A.X protein accumulation and H2A.X incorporation into the chromatin for $\gamma$H2A.X foci formation [322]. These findings suggest that *de novo* H2A.X incorporation into the chromatin is required for efficient DNA damage repair.

In HeLa cells, DNA damage induction by etoposide has no effect on H2A.X poly(A)$^+$ mRNA levels but does lead to increased H2A.X protein levels in HeLa cells (Figure 4.3A, compare siLuc with and without etoposide, and Figure 5.6A, compare wild-type cells with and without etoposide). This is in agreement with H2A.X protein degradation inhibition as previously suggested [322]. Nevertheless, it is possible that the increase in H2A.X protein level upon DNA damage is not only caused by the inhibition of H2A.X protein degradation but also by stimulated mRNA translation, as suggested by the slightly increased levels of H2A.X poly(A)$^+$ mRNA in polysome fractions (Figure 4.7). Interestingly, HCT116 cells do not show such an upregulation of the H2A.X protein upon etoposide treatment (Figure 5.6A). Therefore, it is possible, that the dependence on "de novo" H2A.X accumulation for efficient $\gamma$H2A.X foci formation is increased in HeLa cells but less important in HCT116 cells, as further discussed in Section 7.4.

In HeLa cells, the depletion of the H2A.X poly(A)$^+$ mRNA leads to decreased $\gamma$H2A.X signalling in both non-S-phase and S-phase cells (Figure 4.3D and E). The latter express the S-phase-specific H2A.X SL mRNA, which upon the 2 hr etoposide treatment remains

largely present in the cytoplasm (Figure 4.6D). However, if the H2A.X SL mRNA was able to generate *de novo* H2A.X protein in the S-phase for efficient γH2A.X signalling, S-phase cells would show increased γH2A.X signal, especially upon depletion of the H2A.X poly(A)$^+$ mRNA. The fact that this is not the case (Figure 4.3D and E) suggests that the H2A.X SL mRNA is not translated in S-phase cells treated with etoposide. This in turn suggests that the two H2A.X mRNA isoforms are translated by different mechanisms, as the H2A.X poly(A)$^+$ mRNA clearly is able to produce H2A.X protein under DNA damage treatment (Figure 4.3D, see γH2A.X signal in siLuc cells).

Translated RD SL mRNA was shown to bind the CBC at the 5'-end and SLBP at the 3' SL [182]. Communication between 5'- and 3'-ends is mediated by CTIF (Figure 1.6). DNA replication stress was shown to disrupt this interaction eventually leading to mRNA degradation [182]. This translation mechanism most likely also applies to the H2A.X SL mRNA as it is post-transcriptionally regulated similarly to canonical RD SL mRNA [256]. However, it is unknown which translation mechanism governs the H2A.X poly(A)$^+$ mRNA isoform. As it contains the same sequence elements than H2A.X SL mRNA, it may also be translated in a CBC-CTIF-SLBP-dependent manner. However, as mentioned above, siLuc control cells treated with etoposide efficiently form γH2A.X foci in both non-S-phase and S-phase cells (Figure 4.3), thus it is unlikely that H2A.X poly(A)$^+$ and SL mRNA isoforms are translated by the same mechanism. It rather suggests that H2A.X poly(A)$^+$ mRNA is translated by the same mechanism as "bulk" polyadenylated mRNA, where the CBC is substituted by eIF4E after the pioneer rounds of translation and 5'-3'-communication occurs through eIF4E, eIF4G and PABP [74] (Figures 1.6 and 7.1). In this scenario, DNA damaging agents would only inhibit translation of RD histone SL mRNA, including H2A.X SL mRNA, leaving the translation of polyadenylated mRNA unaffected, such as H2A.X poly(A)$^+$ mRNA (Figure 7.1). Thus the first step affecting RD histone mRNA following DNA damage induction is the inhibition of translation. RD histone mRNA decay is known to be relatively rapid, however, it is also known that H2AFX transcription is not as strongly inhibited by DNA replication stress as canonical RD histone genes [256]. It is therefore possible that 2 hr etoposide treatment leads to the inhibition of H2A.X SL mRNA translation but due to continued production of H2A.X SL mRNA, the overall decrease in H2A.X SL mRNA levels is less rapid. This may explain why S-phase cells also show reduced γH2A.X signalling (decreased H2A.X protein production) even though

the H2A.X SL mRNA is still present. Nevertheless, it remains to be determined if such a differential translation mechanism exists. The polysome gradient experiment shows somewhat increased H2A.X poly(A)$^+$ mRNA in the polysome fractions compared to its SL mRNA counterpart. However, it is unclear whether this is a result of decreased levels of H2A.X SL mRNA or increased H2A.X poly(A)$^+$ mRNA translation (Figure 4.7).

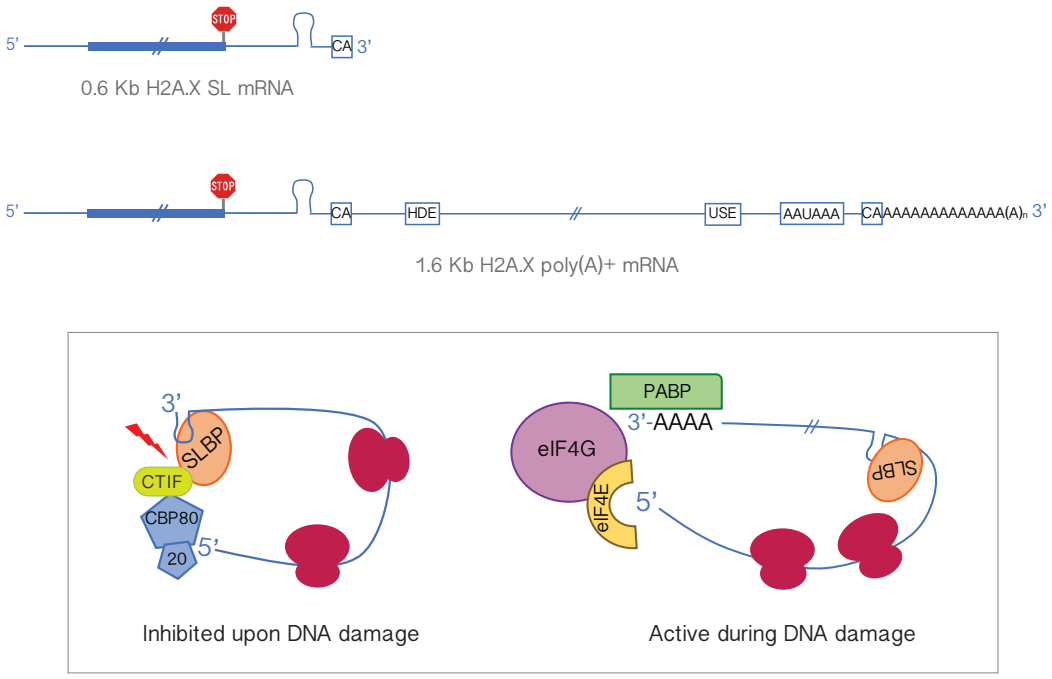



**Figure 7.1 Proposed model for alternative H2A.X mRNA translation mechanisms.** The illustrations on the top represent the two mature H2A.X mRNA isoforms. The box shows their respective proposed translation mechanisms. Left: H2A.X SL mRNA translation. Right: H2A.X poly(A)$^+$ mRNA translation (after the pionieer rounds of translation). The red lightening symbol indicates disruption of SLBP-CTIF interaction upon DNA damage induction, as previously suggested [182].

Further study is required to test whether the H2A.X mRNA isoforms are indeed translated by different mechanisms and how important *de novo* H2A.X protein incorporation at the DSB site really is in dividing cells. Non-dividing cells contain much less H2A.X protein compared to actively proliferating cells [311, 328, 329]. Thus they might depend much more on *de novo* H2A.X production than dividing cells that incorporate large amounts of H2A.X into the chromatin during DNA replication. This will be further discussed below.

## 7.4 The H2A.X protein incorporated into the chromatin in S-phase is required for DNA damage signalling

The recent advance in the CRISPR-Cas9 technology has for the first time allowed analysis of H2A.X expression in the absence of the H2A.X SL mRNA by targeted deletion of the histone downstream element (HDE) that abolishes U7-dependent processing (Figure 5.1). HeLa and HCT116 wild-type cells contain different amounts of H2A.X SL mRNA compared to its polyadenylated counterpart (Figure 5.3A). Compared to HeLa cells, HCT116 cells produce much higher levels of H2A.X SL mRNA in S-phase (Figure 5.4). The extent of H2A.X protein depletion in H2A.X SL mRNA "knock out" (dHDE) cells inversely correlates with the levels of this mRNA isoform in the wild-type cells. *I. e.* HCT116 dHDE cells show a higher reduction in H2A.X protein levels compared to HeLa dHDE cells (Figure 5.6A, D and E).

Strikingly, the γH2A.X signal in dHDE cells was also decreased in both S-phase and non-S-phase cells (Figure 5.8). This suggests that it is not solely *de novo* but also existing H2A.X that is already in chromatin that gives rise to γH2A.X foci. This is especially true for HCT116 cells (Figure 5.8). The high levels of H2A.X SL mRNA in these cells may allow incorporation of large amounts of H2A.X protein into the chromatin during DNA replication. Unlike HeLa cells, no accumulation of total H2A.X protein is observed upon DNA damage in HCT116 cells (Figure 5.6A, also compare B and C) and the depletion of the H2A.X poly(A)$^+$ mRNA has no impact on γH2A.X signalling (Supplementary Figure S9), suggesting that in HCT116 cells the H2A.X protein phosphorylated to form γH2A.X is already present in the chromatin.

The above finding also suggests that the decreased γH2A.X signal observed in HeLa cells depleted of the H2A.X poly(A)$^+$ mRNA (Figure 4.3) may be due to a combination of decreased *de novo* H2A.X protein production from the H2A.X poly(A)$^+$ mRNA and decreased H2A.X incorporation into chromatin during S-phase. As HeLa cells have a relatively high level of H2A.X poly(A)$^+$ mRNA in S-phase compared to the H2A.X SL mRNA, it is possible that the H2A.X protein production from the H2A.X poly(A)$^+$ mRNA is required for sufficient H2A.X incorporation into the chromatin in S-phase cells. As the H2A.X poly(A)$^+$ mRNA depletion with siRNAs was performed for 48 or 72 hr and the

mRNA is already depleted after 24 hr, the cells go through one or two rounds of DNA replication during this time period. This may lead to overall less H2A.X deposition and thus to chromatin that contains a lower percentage of H2A.X protein.

Taken together, in dividing cells, the importance of the two H2A.X mRNA isoforms depends on their relative levels. HeLa cells express less SL than poly(A)$^+$ H2A.X mRNA compared to HCT116 cells (Figure 5.3). Therefore, the depletion of the H2A.X poly(A)$^+$ mRNA has a more dramatic effect on HeLa cells (compare Figure 4.3 and Supplementary Figure S9). In contrast, the depletion of the H2A.X SL mRNA has a higher impact on HCT116 cells (compare Figures 5.7 and 5.8, and see summary Figure 7.2). Interestingly, HeLa cells do accumulate *de novo* H2A.X upon DNA damage induction (Figure 5.6A, compare wild-type HeLa cells with and without etoposide), mainly generated from the H2A.X poly(A)$^+$ mRNA, as also suggested by the increased loading of the H2A.X poly(A)$^+$ mRNA onto polysomes upon DNA damage induction compared to the H2A.X SL mRNA (Figure 4.7). In contrast, H2A.X levels remain unchanged in DNA damage conditions in HCT116 cells (Figure 5.6A, compare wild-type HCT116 cells with and without etoposide), which seem to mainly depend on the H2A.X incorporated into chromatin during S-phase, generated mainly by the highly abundant H2A.X SL mRNA. However, given that the H2A.X poly(A)$^+$ mRNA is the only H2A.X mRNA remaining in both HeLa and HCT116 dHDE cells (H2A.X SL mRNA "knock out"), in these cells it is the H2A.X poly(A)$^+$ mRNA isoform that generates the H2A.X protein incorporated during DNA replication (see summary Figure 7.2). It remains to be seen how important the *de novo* H2A.X protein production from the H2A.X poly(A)$^+$ mRNA outside of S-phase is, especially in HCT116 cells. It has been shown that H2A.X enriched promoters decrease the expression of certain genes in HCT116 cells and that its removal from these nucleosomes leads to transcriptional activation [298]. Thus, the H2A.X poly(A)$^+$ mRNA may be required for H2A.X replacement outside of S-phase to regulate the chromatin environment.

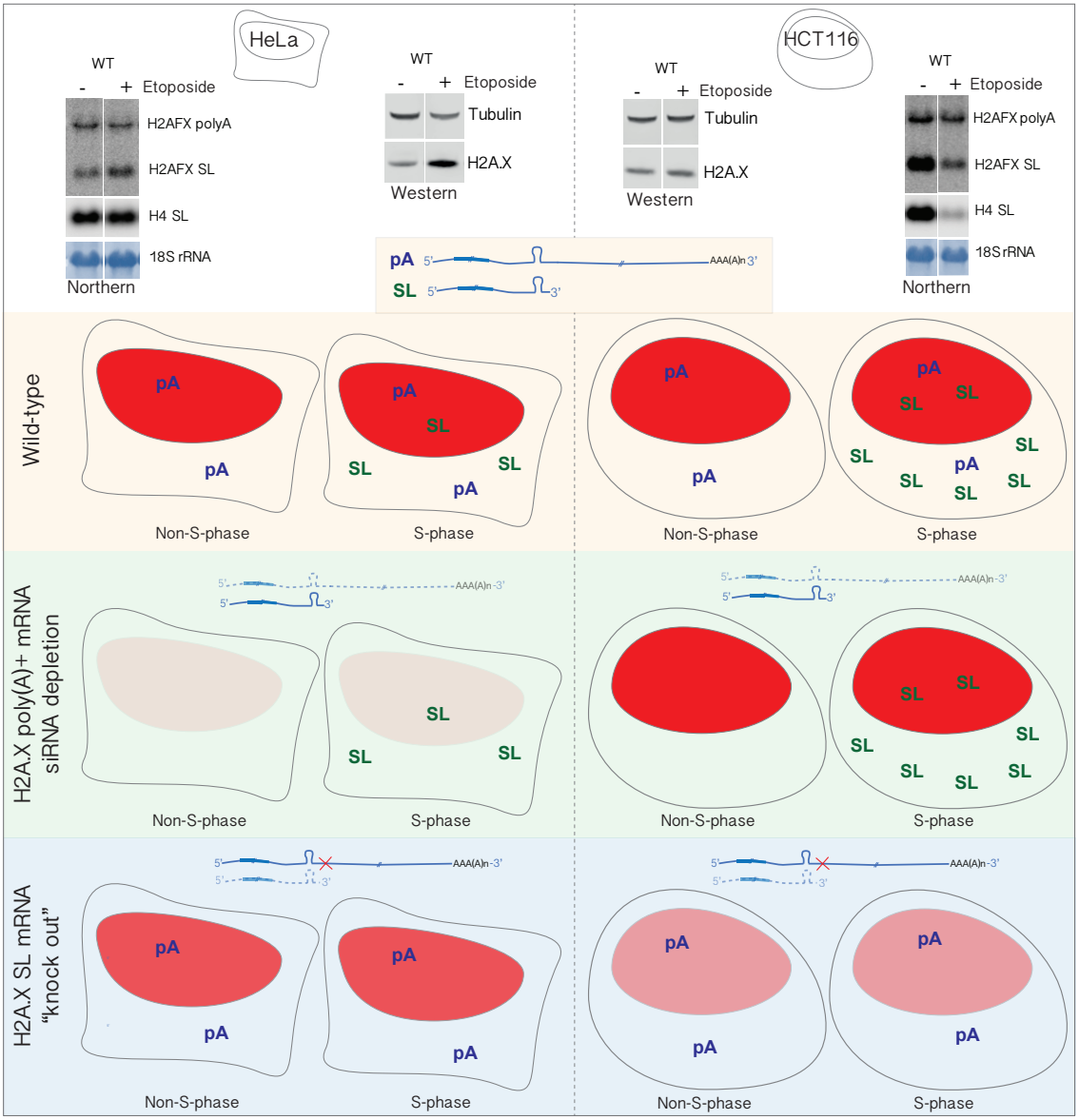



**Figure 7.2 Schematic summarising the effect of the depletion of either of the two H2A.X mRNA isoforms on γH2A.X signalling.** Left panels: HeLa cells. Right panels: HCT116 cells. The northern blot images represent H2A.X mRNA levels in wild-type cells without (-) and with (+) etoposide (taken from Figure 5.5). The western blot images show total H2A.X protein levels in wild-type cells without (-) and with (+) etoposide treatment (taken from Figure 5.6A). 5 µM etoposide were applied for 2 hr before harvesting. Note the accumulation of H2A.X protein upon etoposide treatment in HeLa cells compared to the nearly unchanged levels in HCT116 cells. The illustrations below show the nuclear γH2A.X signal observed in wild-type (yellow box), H2A.X poly(A)$^+$ mRNA depleted (green box) or H2A.X SL mRNA "knock out" cells (blue box). The intensity of the red signal in the nucleus illustrates the γH2A.X signal observed in each condition by immunofluorescence. pA: H2A.X poly(A)$^+$ mRNA (dark blue). SL: H2A.X SL mRNA (dark green). HeLa cells express less H2A.X SL mRNA and thus depend much more on the H2A.X poly(A)$^+$ mRNA for γH2A.X signalling. The γH2A.X signal in HeLa cells likely results from a combination of *de novo* H2A.X protein and H2A.X incorporated in the S-phase by the H2A.X SL mRNA and potentially also by the H2A.X poly(A)$^+$ mRNA. HCT116 cells express much higher levels of H2A.X SL mRNA compared to HeLa cells. Therefore, HCT116 cells rely much more on the H2A.X SL mRNA and the H2A.X protein incorporated into the chromatin in S-phase, mainly by this mRNA isoform as the depletion of the H2A.X poly(A)$^+$ mRNA has no effect on the γH2A.X in these cells. The red cross at the H2A.X poly(A)$^+$ mRNA indicates HDE deletion.

## 7.5 The SL may regulate H2A.X poly(A)$^+$ mRNA levels in S-phase and a conserved region close to the PAS is required for H2A.X poly(A)$^+$ mRNA expression

H2AFX is the only replication-independent histone gene that contains the SL structure typical of RD histone genes. Interestingly, SLBP was not only shown to bind to the SL of both H2A.X SL and poly(A)$^+$ mRNA isoforms, but also to a conserved region upstream of the H2A.X PAS [174]. SLBP specifically recognises the SL [167] and was shown to exclusively bind RD histone mRNA [174]. As the conserved region in the H2A.X poly(A)$^+$ mRNA differs in sequence and structure compared to the SL, its binding by SLBP presumably is indirect, established through another RNA-binding protein (hereinafter referred to as "protein X") [174].

SLBP plays an essential role in the life of RD histone mRNA, influencing pre-mRNA processing, mRNA export, translation and mRNA decay [167, 179–182]. However, which role the SL and SLBP play in the H2A.X poly(A)$^+$ mRNA is unknown. As observed for RD histone mRNA [167, 168, 176], the deletion of the HDE but also of the SL abolish the formation of H2A.X SL mRNA (Figure 6.4). Both these structures are thus required for U7-dependent processing at the H2AFX SL processing site. Regarding the function of the SL in the H2A.X poly(A)$^+$ mRNA, similarly to the deletion of the HDE, the SL deletion leads to somewhat increased H2A.X poly(A)$^+$ mRNA levels as the cells enter S-phase (Figures 6.5, 6.6 and 6.7, compare cell lines #2 and #3 with #1, blue boxes). However, the deletion of the SL allows higher H2A.X poly(A)$^+$ mRNA accumulation in S-phase compared to the HDE deletion (Figures 6.8, 6.5 and 6.6, red boxes; the 5 hr time point corresponds to 90% of cells in S-phase as shown in Supplementary Figure S12). This result suggests that the SL, and presumably SLBP, are involved in controlling H2A.X poly(A)$^+$ mRNA levels or synthesis in S-phase. A possible mechanism for such a regulation is described below.

During S-phase, SLBP levels are high. Therefore, it is possible that it immediately binds the SL as the newly synthesised pre-mRNA emerges from the Pol II exit channel. If the U7-dependent processing reaction does not take place (*e.g.* in the HDE deletion mutant), SLBP "signals" the misprocessing and triggers the degradation of the nascent

mRNA in the nucleus preventing H2A.X poly(A)$^+$ mRNA formation (see proposed model in Figure 7.3). In contrast, in the SL deletion mutant SLBP is unable to bind the nascent mRNA and transcription continues resulting in formation of H2A.X poly(A)$^+$ mRNA. Thus more H2A.X poly(A)$^+$ mRNA will be formed in the SL deletion mutant compared to the HDE deletion mutant. This model would also explain the only mild increase of H2A.X poly(A)$^+$ mRNA levels in S-phase dHDE cells. However, the existence of such a regulatory mechanism remains to be determined.

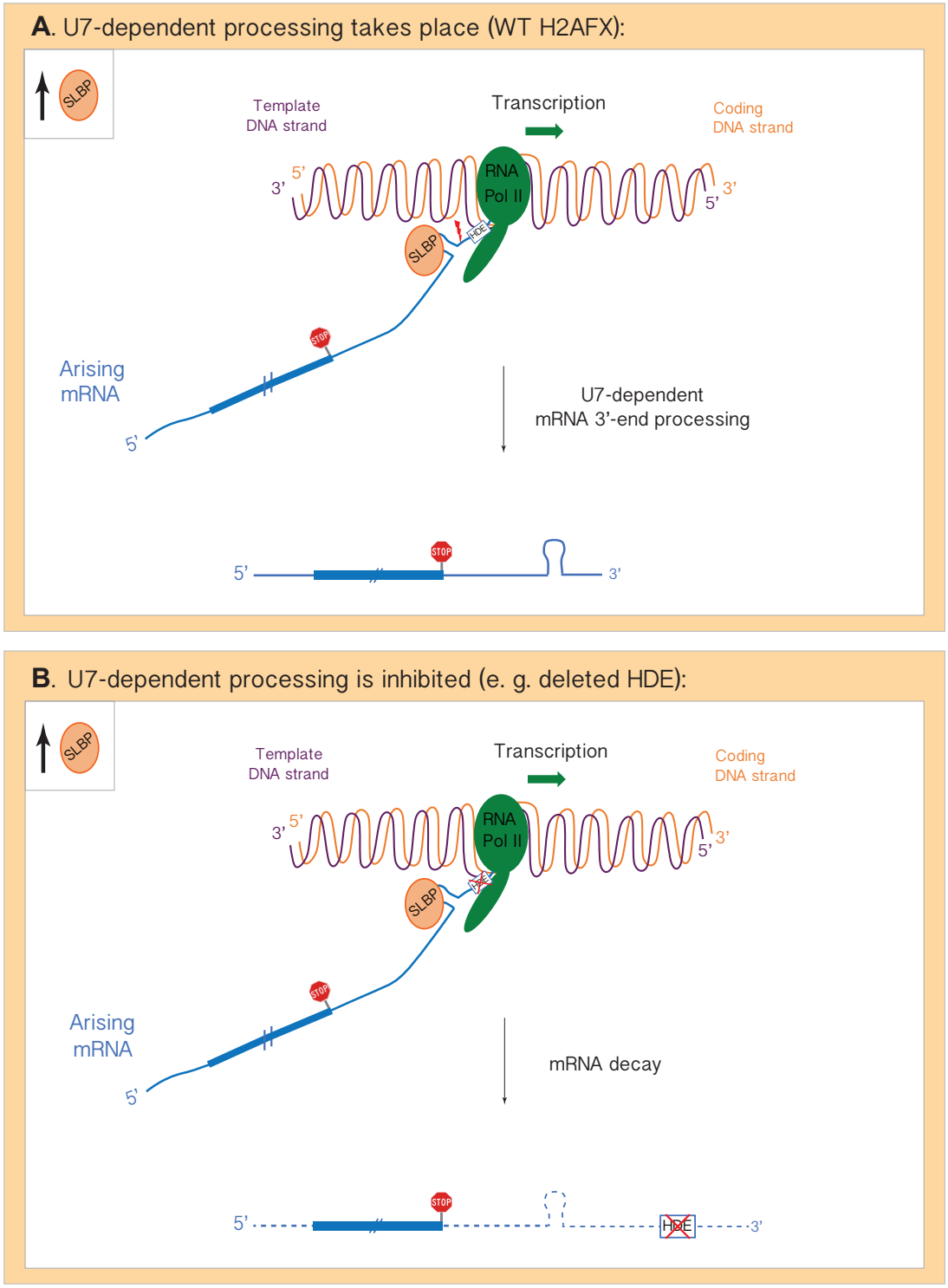



**Figure 7.3 Proposed model for H2A.X poly(A)$^+$ mRNA regulation by SLBP in S-phase.** A) Transcription of wild-type H2AFX in S-phase. U7-dependent processing takes place and mature H2A.X SL mRNA is generated. B) Transcription of H2AFX in a situation where U7-dependent processing is inefficient (*e.g.* HDE deletion). SLBP binds to the SL immediately as it emerges from the Pol II exit channel, "senses" the misprocessing and triggers mRNA degradation. The red cross indicates HDE deletion.

The conservation of the region 5' of the H2AFX PAS throughout vertebrates strongly suggests that this region has a function. Interestingly, it seems to be involved in maintaining H2A.X poly(A)$^+$ mRNA levels as its deletion leads to decreased accumulation of H2A.X poly(A)$^+$ mRNA (Figure 6.3, light green arrows). Such regulation could occur in a co-transcriptional or post-transcriptional manner. If the regulation occurs co-transcriptionally, it could be that an uncharacterised protein binding to the conserved region (protein X) recruits polyadenylation factors and thereby stimulates the cleavage and polyadenylation reaction. SLBP is also known to interact with the conserved region [174], presumably through interaction with protein X. Thus it is possible that in situations where U7-dependent processing does not take place (*e.g.* HDE deletion) Pol II continues transcribing until it encounters the PAS. There is then an interaction between SLBP and protein X that prevents the recruitment of polyadenylation factors and/or stimulates the recruitment of the RNA degradation machinery, ultimately leading to the degradation of the H2A.X poly(A)$^+$ mRNA in S-phase (Figure 7.4A). In contrast, outside of S-phase, when SLBP is low, such interaction would not take place and protein X would not be inhibited by SLBP and would then somehow promote H2A.X poly(A)$^+$ mRNA accumulation (Figure 7.4B). Both these hypothetical co- and post-transcriptional mechanisms are not mutually exclusive and could act simultaneously. The conserved region could also be involved in other steps in the life of the H2A.X poly(A)$^+$ mRNA, such as mRNA export and/or translation, which may be regulated differently upon DNA damage.

Alternatively, if the regulation occurs at post-transcriptional level, protein X might protect the H2A.X poly(A)$^+$ mRNA from degradation. It is unknown if SLBP binds the SL of the H2A.X poly(A)$^+$ mRNA also outside of S-phase. If SLBP only binds to the SL in S-phase, it is possible that SLBP-protein X interaction stimulates H2A.X poly(A)$^+$ mRNA degradation (Figure 7.5A). Whereas, outside of S-phase, when SLBP levels are low, it would not bind to the SL of the H2A.X poly(A)$^+$ mRNA and protein X would not interact with SLBP but likely with other factors favouring the stabilisation of the H2A.X poly(A)$^+$ mRNA (Figure 7.5B).

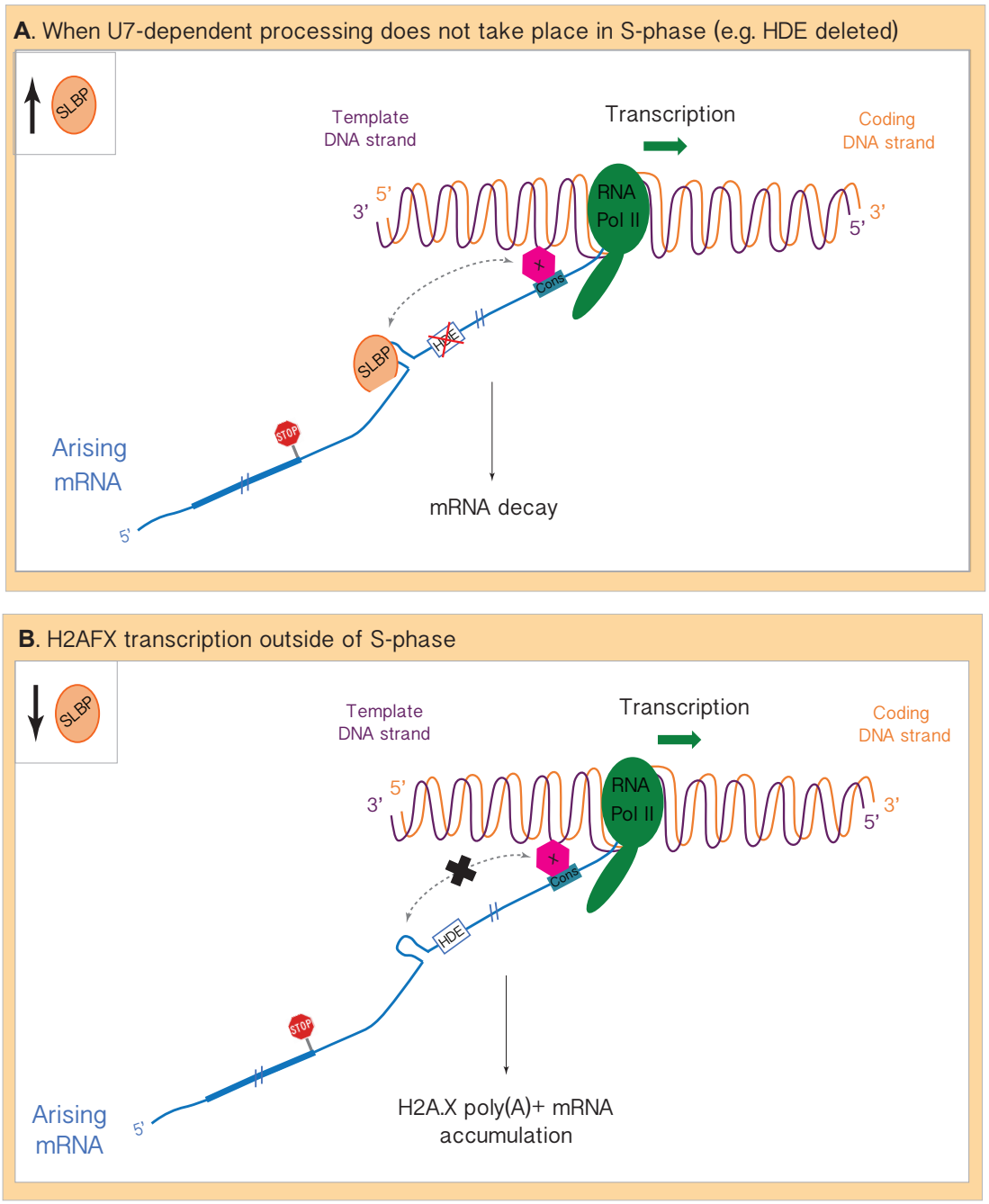



**Figure 7.4 Proposed model for co-transcriptional regulation of the H2A.X poly(A)$^+$ mRNA by the conserved region.** A) Transcription of H2AFX in a situation where U7-dependent processing is inefficient (*e.g.* HDE deletion) in S-phase. SLBP-protein X interaction leads to mRNA decay. B) Outside of S-phase, SLBP levels are low. No interaction between SLBP and protein X is established and H2A.X poly(A)$^+$ mRNA accumulates. The red cross indicates HDE deletion. The black cross indicates that the disruption of the interaction of the SL and the conserved region.

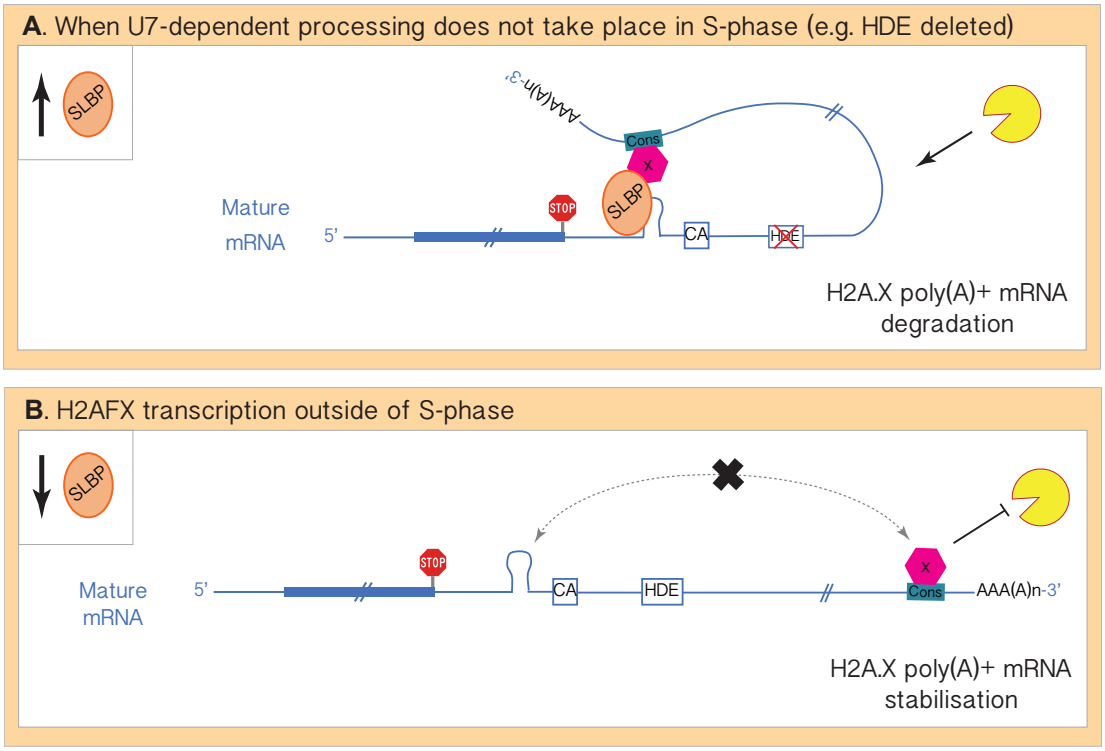



**Figure 7.5 Proposed model for post-transcriptional regulation of the H2A.X poly(A)$^+$ mRNA by the conserved region.** A) H2A.X poly(A)$^+$ mRNA generated in S-phase as a result of inefficient U7-dependent processing (*e.g.* HDE deletion). The SLBP-protein X interaction triggers post-transcriptional mRNA decay. B) Outside of S-phase, SLBP does not bind the SL of the H2A.X poly(A)$^+$ mRNA, no interaction between SLBP and protein X is established and the latter executes its mRNA stabilising function. The red cross indicates HDE deletion. The black cross indicates that the disruption of the interaction of the SL and the conserved region.

## 7.6 General conclusions

Replication-dependent histone genes have been shown to express polyadenylated mRNA in differentiated cells, as a source for replacement histones. However, my results show that the formation of read-through polyadenylated histone mRNA from RD histone genes is not only a result of the differentiation program, as highly proliferating HeLa cells also express some of these mRNA isoforms. Furthermore, these polyadenylated histone mRNA isoforms in HeLa cells were also detected in the cytoplasm making them potential translation substrates.

H2AFX is the best known histone gene that expresses both SL and poly(A)$^+$ mRNA isoforms, which have replication-dependent and replication-independent characteristics, respectively. My results show that the H2A.X poly(A)$^+$ mRNA is mainly generated by H2AFX transcription outside of the S-phase. A minor fraction may be produced in S-phase as a result of inefficient U7-dependent processing. Furthermore, H2A.X poly(A)$^+$ mRNA levels are upregulated in G2/M in both HeLa and HCT116 cells. Additionally, unlike RD histone genes, whose expression is tighly controlled by components of the histone locus body (HLB) associated around the histone gene cluster, the expression of the S-phase-specific H2A.X SL mRNA mainly seems to depend on the presence of the histone 3'-end processing factors and is not controlled by factors that restrict RD histone gene expression in the HLB.

By depleting either of these two mRNA isoforms, I show that their requirement for efficient DNA damage signalling depends on their relative levels. Thus cells that express relatively low levels of the H2A.X SL mRNA, such as HeLa and RPE-1 cells, mainly depend on the H2A.X poly(A)$^+$ mRNA, for *de novo* H2A.X protein production outside of S-phase but possibly also for sufficient H2A.X incorporation into newly synthesised chromatin during S-phase. In contrast, cells that express high levels of the H2A.X SL mRNA, such as HCT116 cells, mainly depend on this mRNA isoform.

The SL structure in the H2A.X poly(A)$^+$ mRNA seems to have a role in restricting its accumulation in S-phase, possibly through an interaction with the conserved region upstream of the polyadenylation site. It remains to be determined if these regulatory effects occur co- or post-transcriptionally.

## 7.7 Future directions

Different cell lines express different relative levels of the two H2A.X mRNA isoforms. To test whether this is caused by a difference in the levels of 3'-end processing machinery components, U7 snRNA northern blots as well as western blotting for histone 3'-end processing factors including SLBP, FLASH and LSM11 will be performed.

Next, experiments will be performed to test whether the H2A.X protein accumulation upon DNA damage induction in HeLa cells is due to a change in protein stability (using proteasome inhibitors) or translation (performing polysome gradients followed by northern blotting). Furthermore, *de novo* H2A.X incorporation into nucleosomes on chromatin will be analysed in the different phases of the cell cycle, in control and DNA damage situations, using tagged H2A.X protein (*e.g.* Halo tag) and live cell imaging. This could be complemented with $^{35}$S labelling followed by H2A.X chromatin immunoprecipitation (ChIP).

Additionally, the translation mechanism of the two H2A.X mRNA isoforms will be assessed using an eIF4E-specific inhibitor (4EGI-1) and polysome gradient fractionation.

The importance of the SL and the conserved region will be further studied using different HeLa Flp-In T-REx constructs. A very informative construct will be an siRNA-resistant SL deletion construct. Such a construct would allow depletion of the endogenous H2A.X poly(A)$^+$ mRNA whereas it would show if the levels of the transgene H2A.X poly(A)$^+$ mRNA change in the absence of the SL structure. Furthermore, a H2AFX transgene deleted of both the SL and the conserved region will be used to test their synergistic effect. Finally, the *trans*-acting factor recognising the conserved region upstream of the H2A.X polyadenulation site (protein X) will be identified and its regulatory function on the H2A.X poly(A)$^+$ mRNA levels will be assessed.

Given that the majority of experiments were performed on cancer cells (HeLa and HCT116), an open question that remains in the field is the significance of the two H2A.X mRNA isoforms in untransformed cells and in the entire organism. Especially considering that only vertebrate H2AFX genes have evolved to contain the SL structure. Therefore, I propose that the generation of transgenic animals deleted of either the H2AFX SL structure (H2A.X SL mRNA "knock out") or the polyadenylation site (H2A.X poly(A)$^+$ mRNA

"knock out") would be an elegant solution to study the systematic relevance of these two H2A.X mRNA isoforms.

# References


[1] N. J. Proudfoot. Transcriptional termination in mammals: Stopping the RNA polymerase II juggernaut. *Science*, 352(6291):aad9926, 2016.

[2] R. G. Roeder and W. J. Rutter. Multiple Forms of DNA-dependent RNA Polymerase in Eukaryotic Organisms. *Nature*, 224(5216):234–237, 1969.

[3] C. Engel, T. Gubbey, S. Neyer, S. Sainsbury, C. Oberthuer, C. Baejen, C. Bernecky, and P. Cramer. Structural Basis of RNA Polymerase I Transcription Initiation. *Cell*, 169(1):120–131.e22, 2017.

[4] T. Moss, F. Langlois, T. Gagnon-Kugler, and V. Stefanovsky. A housekeeper with power of attorney: The rRNA genes in ribosome biogenesis. *Cellular and Molecular Life Sciences*, 64(1):29–49, 2007.

[5] Guillermo Abascal-Palacios, Ewan Phillip Ramsay, Fabienne Beuron, Edward Morris, and Alessandro Vannini. Structural basis of RNA polymerase III transcription initiation. *Nature*, 553:301–306, 2018.

[6] T. I. Lee and . A. Young. Transcription of Eukaryotic Protein-Coding Genes. *Annual Review of Genetics*, 34:77–137, 2000.

[7] S. Sainsbury, C. Bernecky, and P. Cramer. Structural basis of transcription initiation by RNA polymerase II. *Nature Reviews Molecular Cell Biology*, 16(3):129–143, 2015.

[8] I. Jonkers and J. T. Lis. Getting up to speed with transcription elongation by RNA polymerase II. *Nature Reviews Molecular Cell Biology*, 16(3):167–177, 2015.

[9] F. X. Chen, E. R. Smith, and A. Shilatifard. Born to run: control of transcription elongation by RNA polymerase II. *Nature Reviews Molecular Cell Biology*, 19(7):464–478, 2018.

[10] A. C. M. Cheung and P. Cramer. A movie of RNA polymerase II transcription. *Cell*, 149(7):1431–7, 2012.

[11] K. Adelman and J. T. Lis. Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nature Reviews Genetics*, 13(10):720–731, 2012.

[12] A. R. Krebs, D. Imanci, L. Hoerner, D. Gaidatzis, L. Burger, and D. Schübeler. Genome-wide Single-Molecule Footprinting Reveals High RNA Polymerase II Turnover at Paused Promoters. *Molecular Cell*, 67(3):411–422.e4, 2017.

[13] F. W. Martinez-Rucobo, R. Kohler, M. van de Waterbeemd, A. J. R. Heck, M. Hemann, F. Herzog, H. Stark, and P. Cramer. Molecular Basis of Transcription-Coupled Pre-mRNA Capping. *Molecular Cell*, 58(6):1079–1089, 2015.

[14] J. D. Lewis and E. Izaurralde. The role of the cap structure in RNA processing and nuclear export. *European Journal of Biochemistry*, 247(2):461–9, 1997.

[15] K. Inoue, M. Ohno, H. Sakamoto, and Y. Shimura. Effect of the cap structure on pre-mRNA splicing in Xenopus oocyte nuclei. *Genes and Development*, 3(9):1472–9, 1989.

[16] M. Pabis, N. Neufeld, M. C. Steiner, T. Bojic, Y. Shav-Tal, and K. M. Neugebauer. The nuclear cap-binding complex interacts with the U4/U6·U5 tri-snRNP and promotes spliceosome assembly in mammalian cells. *RNA*, 19(8):1054–63, 2013.

[17] S. M. Flaherty, P. Fortes, E. Izaurralde, I. W. Mattaj, and G. M. Gilmartin. Participation of the nuclear cap binding complex in pre-mRNA 3' processing. *Proceedings of the National Academy of Sciences of the United States of America*, 94(22):11893–8, 1997.

[18] E. Izaurralde and S. Adam. Transport of macromolecules between the nucleus and the cytoplasm. *RNA*, 4(4):351–64, 1998.

[19] T. Nojima, T. Hirose, H. Kimura, and M. Hagiwara. The interaction between cap-binding complex and RNA export factor is required for intronless mRNA export. *The Journal of Biological Chemistry*, 282(21):15645–51, 2007.

[20] L. E. Maquat, W. Tarn, and O. Isken. The pioneer round of translation: features and functions. *Cell*, 142(3):368–74, 2010.

[21] C. A. Beelman and R. Parker. Degradation of mRNA in Eukaryotes Review. *Cell*, 81(2):179–188, 1995.

[22] T. Nojima, K. Rebelo, T. Gomes, A. R. Grosso, N. J. Proudfoot, and M. Carmo-Fonseca. RNA Polymerase II Phosphorylated on CTD Serine 5 Interacts with the Spliceosome during Co-transcriptional Splicing. *Molecular Cell*, 72(2):369–379.e4, 2018.

[23] A. A. Patel and J. A. Steitz. Splicing double: insights from the second spliceosome. *Nature Reviews Molecular Cell Biology*, 4(12):960–970, 2003.

[24] M. C. Wahl, C. L. Will, and R. Lührmann. Leading Edge The Spliceosome: Design Principles of a Dynamic RNP Machine. *Cell*, 136(4):701–18, 2009.

[25] A. Nott, H. Le Hir, and M. J. Moore. Splicing enhances translation in mammalian cells: an additional function of the exon junction complex. *Genes and Development development*, 18(2):210–22, 2004.

[26] S. Millevoi, C. Loulergue, S. Dettwiler, S. Z. Karaa, W. Keller, M. Antoniou, S. Vagner, P. Bork, W. Keller, M. Wilm, M. W. Hentze, and A. E. Kulozik. An interaction between U2AF 65 and CF Im links the splicing and 3' end processing machineries. *The EMBO Journal*, 25(20):4854–4864, 2006.

[27] U. Braunschweig, S. Gueroussov, A. M. Plocik, B. R. Graveley, and B. J. Blencowe. Dynamic integration of splicing within gene regulatory pathways. *Cell*, 152(6):1252–69, 2013.

[28] M. Müller-McNicoll, V. Botti, A. M. de Jesus Domingues, H. Brandl, O. D. Schwich, M. C. Steiner, T. Curk, I. Poser, K. Zarnack, and K. M. Neugebauer. SR proteins are NXF1 adaptors that link alternative RNA processing to mRNA export. *Genes and Development*, 30(5):553–66, 2016.

[29] B. Roy, L. M. Haupt, and L. R. Griffiths. Review: Alternative Splicing (AS) of Genes As An Approach for Generating Protein Complexity. *Current Genomics*, 14(3):182–94, 2013.

[30] F. E. Baralle and J. Giudice. Alternative splicing as a regulator of development and tissue identity. *Nature Reviews Molecular Cell Biology*, 18(7):437–451, 2017.

[31] E. Whitelaw and N. Proudfoot. Alpha-thalassaemia caused by a poly(A) site mutation reveals that transcriptional termination is linked to 3' end processing in the human alpha 2 globin gene. *The EMBO Journal*, 5(11):2915–2922, 1986.

[32] J. Zhao, L. Hyman, and C. Moore. Formation of mRNA 3' Ends in Eukaryotes: Mechanism, Regulation, and Interrelationships with Other Steps in mRNA Synthesis. *Microbiology and Molecular Biology Reviews*, 63(2):405–445, 1999.

[33] T. Nojima, M. Carmo-Fonseca, N. J. Proudfoot, T. S. Gomes, A. R. Fialho Grosso, H. Kimura, M. J. Dye, and S. Dhir. Mammalian NET-Seq Reveals Genome-wide Nascent Transcription Coupled to RNA Processing. *Cell*, 161(3):526–540, 2015.

[34] K. Skourti-Stathaki, N. J. Proudfoot, and N. Gromak. Human Senataxin Resolves RNA/DNA Hybrids Formed at Transcriptional Pause Sites to Promote Xrn2-Dependent Termination. *Molecular Cell*, 42(6):794–805, 2011.

[35] J. D. Eaton, L. Davidson, D. L. V. Bauer, T. Natsume, M. T. Kanemaki, and S. West. Xrn2 accelerates termination by RNA polymerase II, which is underpinned by CPSF73 activity. *Genes and Development*, 32(2):127–139, 2018.

[36] J. Park, M. Kang, and M. Kim. Unraveling the mechanistic features of RNA polymerase II termination by the 5'-3' exoribonuclease Rat1. *Nucleic Acids Research*, 43(5):2625–37, 2015.

[37] N. Fong, K. Brannan, B. Erickson, H. Kim, M. A. Cortazar, R. M. Sheridan, T. Nguyen, S. Karp, and D. L. Bentley. Effects of Transcription Elongation Rate and Xrn2 Exonuclease Activity on RNA Polymerase II Termination Suggest Widespread Kinetic Competition. *Molecular Cell*, 60(2):256–267, 2015.

[38] Z. Dominski, X. Yang, and W. F. Marzluff. The Polyadenylation Factor CPSF-73 Is Involved in Histone-Pre-mRNA Processing. *Cell*, 123(1):37–48, 2005.

[39] X. Yang, K. D. Sullivan, W. F. Marzluff, and Z. Dominski. Studies of the 5' exonuclease and endonuclease activities of CPSF-73 in histone pre-mRNA processing. *Molecular and Cellular Biology*, 29(1):31–42, 2009.

[40] S. T. Smale and R. Tjian. Transcription of herpes simplex virus tk sequences under the control of wild-type and mutant human RNA polymerase I promoters. *Molecular and Cellular Biology*, 5(2):352–62, 1985.

[41] S. S. Sisodia, B. Sollner-Webb, and D. W. Cleveland. Specificity of RNA maturation pathways: RNAs transcribed by RNA polymerase III are not substrates for splicing or polyadenylation. *Molecular and Cellular Biology*, 7(10):3602–12, 1987.

[42] N. J. Proudfoot, A. Furger, and M. J. Dye. Integrating mRNA Processing with Transcription. *Cell*, 108(4):501–512, 2002.

[43] N. J. Proudfoot. New perspectives on connecting messenger RNA 3' end formation to transcription. *Current Opinion in Cell Biology*, 16(3):272–278, 2004.

[44] M. J. Moore and N. J. Proudfoot. Pre-mRNA processing reaches back to transcription and ahead to translation. *Cell*, 136(4):688–700, 2009.

[45] S. Shukla and S. Oberdoerffer. Co-transcriptional regulation of alternative pre-mRNA splicing. *Biochimica et Biophysica Acta*, 1819(7):673–83, 2012.

[46] D. L. Bentley. Coupling mRNA processing with transcription in time and space. *Nature Reviews Genetics*, 15(3):163–75, 2014.

[47] E. C. Merkhofer, P. Hu, and T. L. Johnson. Introduction to cotranscriptional RNA splicing. *Methods in Molecular Biology*, 1126:83–96, 2014.

[48] N. H. Gehring, E. Wahle, and U. Fischer. Deciphering the mRNP Code: RNA-Bound Determinants of Post-Transcriptional Gene Regulation. *Trends in Biochemical Sciences*, 42(5):369–382, 2017.

[49] S. F. Mitchell and R. Parker. Principles and Properties of Eukaryotic mRNPs. *Molecular Cell*, 54(4):547–558, 2014.

[50] M. Müller-McNicoll and K. M. Neugebauer. How cells get the message: dynamic assembly and function of mRNA–protein complexes. *Nature Reviews Genetics*, 14(4):275–287, 2013.

[51] A. Meinhart, T. Kamenski, S. Hoeppner, S. Baumli, and P. Cramer. A structural perspective of CTD function. *Genes and Development*, 19(12):1401–15, 2005.

[52] S. Egloff and S. Murphy. Cracking the RNA polymerase II CTD code. *Trends in Genetics*, 24(6):280–8, 2008.

[53] S. McCracken, N. Fong, E. Rosonina, K. Yankulov, G. Brothers, D. Siderovski, A. Hessel, S. Foster, S. Shuman, and D. L. Bentley. 5'-Capping enzymes are targeted to pre-mRNA by binding to the phosphorylated carboxy-terminal domain of RNA polymerase II. *Genes and Development*, 11(24):3306–18, 1997.

[54] S. C. Schroeder, B. Schwer, S. Shuman, and D. Bentley. Dynamic association of capping enzymes with transcribing RNA polymerase II. *Genes and Development*, 14(19):2435–40, 2000.

[55] N. Fong and D. L. Bentley. Capping, splicing, and 3' processing are independently stimulated by RNA polymerase II: different functions for different segments of the CTD. *Genes and Development*, 15(14):1783–95, 2001.

[56] A. Meinhart and P. Cramer. Recognition of RNA polymerase II carboxy-terminal domain by 3'-RNA-processing factors. *Nature*, 430(6996):223–226, 2004.

[57] Y. Hirose and J. L. Manley. RNA polymerase II and the integration of nuclear events. *Genes and Development*, 14(12):1415–29, 2000.

[58] J. Hsin and J. L. Manley. The RNA polymerase II CTD coordinates transcription and RNA processing. *Genes and Development*, 26(19):2119–37, 2012.

[59] M. Beck and E. Hurt. The nuclear pore complex: understanding its function through structural insight. *Nature Reviews Molecular Cell Biology*, 18(2):73–89, 2017.

[60] D. Mohr, S. Frey, T. Fischer, T. Güttler, and D. Görlich. Characterisation of the passive permeability barrier of nuclear pore complexes. *The EMBO Journal*, 28(17):2541–2553, 2009.

[61] A. K. Friedman and L. A. Baker. Synthetic hydrogel mimics of the nuclear pore complex display selectivity dependent on FG-repeat concentration and electrostatics. *Soft Matter*, 12(47):9477–9484, 2016.

[62] V. O. Wickramasinghe, P. I. A. McMurtrie, A. D. Mills, Y. Takei, S. Penrhyn-Lowe, Y. Amagase, S. Main, J. Marr, M. Stewart, and R. A. Laskey. mRNA export from mammalian cell nuclei is dependent on GANP. *Current Biology*, 20(1):25–31, 2010.

[63] J. Katahira. Nuclear Export of Messenger RNA. *Genes*, 6(2):163–184, 2015.

[64] R. A. Elbarbary and L. E. Maquat. Coupling pre-mRNA splicing and 3' end formation to mRNA export: alternative ways to punch the nuclear export clock. *Genes and Development*, 30(5):487–8, 2016.

[65] M. Soheilypour and M. R. K. Mofrad. Quality control of mRNAs at the entry of the nuclear pore: Cooperation in a complex molecular system. *Nucleus*, 9(1):202–211, 2018.

[66] H. Cheng, K. Dufu, C. Lee, J. L. Hsu, A. Dias, and R. Reed. Human mRNA export machinery recruited to the 5' end of mRNA. *Cell*, 127(7):1389–400, 2006.

[67] R. Reed and E. Hurt. A Conserved mRNA Export Machinery Coupled to pre-mRNA Splicing. *Cell*, 108(4):523–531, 2002.

[68] D. H. Lin, A. R. Correia, S. W. Cai, F. M. Huber, C. A. Jette, and A. Hoelz. Structural and functional analysis of mRNA export regulation by the nuclear pore complex. *Nature Communications*, 9(1):2319, 2018.

[69] J. M. Halstead, T. Lionnet, J. H. Wilbertz, F. Wippich, A. Ephrussi, R. H. Singer, and J. A. Chao. An RNA biosensor for imaging the first round of translation from single cells to living animals. *Science*, 347(6228):1367–1671, 2015.

[70] Y. Ishigaki, X. Li, G. Serin, and L. E. Maquat. Evidence for a Pioneer Round of mRNA Translation: mRNAs Subject to Nonsense-Mediated Decay in Mammalian Cells Are Bound by CBP80 and CBP20. *Cell*, 106(5):607–617, 2001.

[71] O. Isken and L. E. Maquat. The multiple lives of NMD factors: balancing roles in gene and genome regulation. *Nature Reviews Genetics*, 9(9):699–712, 2008.

[72] N. H. Gehring, S. Lamprinaki, A. E. Kulozik, and M. W. Hentze. Disassembly of exon junction complexes by PYM. *Cell*, 137(3):536–48, 2009.

[73] S. Lykke-Andersen and T. H. Jensen. Nonsense-mediated mRNA decay: an intricate machinery that shapes transcriptomes. *Nature Reviews Molecular Cell Biology*, 16(11):665–677, 2015.

[74] H. Sato and L. E. Maquat. Remodeling of the pioneer translation initiation complex involves translation and the karyopherin importin beta. *Genes and Development*, 23(21):2537–50, 2009.

[75] S. H. Kessler and A. B. Sachs. RNA recognition motif 2 of yeast Pab1p is required for its functional interaction with eukaryotic translation initiation factor 4G. *Molecular and Cellular Biology*, 18(1):51–7, 1998.

[76] H. Imataka, A. Gradi, and N. Sonenberg. A newly identified N-terminal amino acid sequence of human eIF4G binds poly(A)-binding protein and functions in poly(A)-dependent translation. *The EMBO Journal*, 17(24):7480–7489, 1998.

[77] R. J. Jackson, C. U. T. Hellen, and T. V. Pestova. The mechanism of eukaryotic translation initiation and principles of its regulation. *Nature Reviews Molecular Cell Biology*, 11(2):113–27, 2010.

[78] J. Costello, L. M. Castelli, W. Rowe, C. J. Kershaw, D. Talavera, S. S. Mohammad-Qureshi, P. F .G. Sims, C. M. Grant, G. D. Pavitt, S. J. Hubbard, and M. P. Ashe. Global mRNA selection mechanisms for translation initiation. *Genome Biology*, 16(1):10, 2015.

[79] S. K. Archer, N. E. Shirokikh, C. V. Hallwirth, T. H. Beilharz, and T. Preiss. Probing the closed-loop model of mRNA translation in living cells. *RNA Biology*, 12(3):248–254, 2015.

[80] X. Wang, W. Xi, S. Toomey, Y. Chiang, J. Hasek, T. M. Laue, and C. L. Denis. Stoichiometry and Change of the mRNA Closed-Loop Factors as Translating Ribosomes Transit from Initiation to Elongation. *PLOS ONE*, 11(3):e0150616, 2016.

[81] Q. Vicens, J. S. Kieft, and O. S. Rissland. Revisiting the Closed-Loop Model and the Nature of mRNA 5'-3' Communication. *Molecular Cell*, 72(5):805–812, 2018.

[82] A. G. Hinnebusch. The Scanning Mechanism of Eukaryotic Translation Initiation. *Annual Review of Biochemistry*, 83(1):779–812, 2014.

[83] V. Ramakrishnan. Ribosome structure and the mechanism of translation. *Cell*, 108(4):557–72, 2002.

[84] A. G. Hinnebusch and J. R. Lorsch. The mechanism of eukaryotic translation initiation: new insights and challenges. *Cold Spring Harbor Perspectives in Biology*, 4(10):a011544, 2012.

[85] A. P. Schuller and R. Green. Roadblocks and resolutions in eukaryotic translation. *Nature Reviews Molecular Cell Biology*, 19(8):526–541, 2018.

[86] G. Hanson and J. Coller. Codon optimality, bias and usage in translation and mRNA decay. *Nature Reviews Molecular Cell Biology*, 19(1):20–30, 2017.

[87] D. E. Andreev, P. B. F. O'Connor, G. Loughran, S. E. Dmitriev, P. V. Baranov, and I. N. Shatsky. Insights into the mechanisms of eukaryotic translation gained with ribosome profiling. *Nucleic Acids Research*, 45(2):513–526, 2017.

[88] C. A. Chen and A. Shyu. Mechanisms of deadenylation-dependent decay. *Wiley Interdisciplinary Reviews: RNA*, 2(2):167–83, 2011.

[89] M. Schlackow, T. Nojima, T. Gomes, A. Dhir, M. Carmo-Fonseca, and N. J. Proudfoot. Distinctive Patterns of Transcription and RNA Processing for Human lincRNAs. *Molecular Cell*, 65(1):25–38, 2017.

[90] M. Song, Y. Li, and M. Kiledjian. Multiple mRNA decapping enzymes in mammalian cells. *Molecular Cell*, 40(3):423–32, 2010.

[91] D. Reines. Decapping goes nuclear. *Molecular Cell*, 46(3):241–2, 2012.

[92] K. Brannan, H. Kim, B. Erickson, K. Glover-Cutter, S. Kim, N. Fong, L. Kiemele, K. Hansen, R. Davis, J. Lykke-Andersen, and D. L. Bentley. mRNA decapping factors and the exonuclease Xrn2 function in widespread premature termination of RNA polymerase II transcription. *Molecular Cell*, 46(3):311–24, 2012.

[93] J. C. Zinder and C. D. Lima. Targeting RNA for processing or destruction by the eukaryotic RNA exosome and its cofactors. *Genes and Development*, 31(2):88–100, 2017.

[94] C. G. Körner and E. Wahle. Poly(A) tail shortening by a mammalian poly(A)-specific 3'-exoribonuclease. *The Journal of Biological Chemistry*, 272(16):10448–56, 1997.

[95] T. Chang, A. Yamashita, C. A. Chen, Y. Yamashita, W. Zhu, S. Durdan, A. Kahvejian, N. Sonenberg, and A. Shyu. UNR, a new partner of poly(A)-binding protein, plays a key role in translationally coupled mRNA turnover mediated by the c-fos major coding-region determinant. *Genes and Development*, 18(16):2010–23, 2004.

[96] A. C. Goldstrohm and M. Wickens. Multifunctional deadenylase complexes diversify mRNA control. *Nature Reviews Molecular Cell Biology*, 9(4):337–344, 2008.

[97] E. Wahle and G. S. Winkler. RNA decay machines: Deadenylation by the Ccr4–Not and Pan2–Pan3 complexes. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, 1829(6-7):561–570, 2013.

[98] C. J. Norbury. Cytoplasmic RNA: a case of the tail wagging the dog. *Nature Reviews Molecular Cell Biology*, 14(10):643–653, 2013.

[99] E. Dehlin, M. Wormington, C. G. Körner, and E. Wahle. Cap-dependent deadenylation of mRNA. *The EMBO Journal*, 19(5):1079–1086, 2000.

[100] S. Meyer, C. Temme, and E. Wahle. Messenger RNA Turnover in Eukaryotes: Pathways and Enzymes. *Critical Reviews in Biochemistry and Molecular Biology*, 39(4):197–216, 2004.

[101] R. Tomecki, M. S. Kristiansen, S. Lykke-Andersen, A. Chlebowski, K. M. Larsen, R. J. Szczesny, K. Drazkowska, A. Pastula, J. S. Andersen, P. P. Stepien, A. Dziembowski, and T. H. Jensen. The human core exosome interacts with differentially localized processive RNases: hDIS3 and hDIS3L. *The EMBO journal*, 29(14):2342–57, 2010.

[102] A. Labno, R. Tomecki, and A. Dziembowski. Cytoplasmic RNA decay pathways - Enzymes and mechanisms. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 1863(12):3125–3147, 2016.

[103] D. D. Scott and C. J. Norbury. RNA decay via 3' uridylation. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, 1829(6-7):654–665, 2013.

[104] M. K. Slevin, S. Meaux, J. D. Welch, R. Bigler, P. L. Miliani de Marval, W. Su, R. E. Rhoads, J. F. Prins, and W. F. Marzluff. Deep sequencing shows multiple oligouridylations are required for 3' to 5' degradation of histone mRNAs on polyribosomes. *Molecular Cell*, 53(6):1020–1030, 2014.

[105] J. Lim, M. Ha, H. Chang, S. C. Kwon, D. K. Simanshu, D. J. Patel, and V. N.. Kim. Uridylation by TUT4 and TUT7 marks mRNA for degradation. *Cell*, 159(6):1365–76, 2014.

[106] H. Chang, J. Lim, M. Ha, and V. N. Kim. TAIL-seq: genome-wide determination of poly(A) tail length and 3' end modifications. *Molecular Cell*, 53(6):1044–1052, 2014.

[107] L. P. Ford, P. S. Bagga, and J. Wilusz. The poly(A) tail inhibits the assembly of a 3'-to-5' exonuclease in an in vitro RNA stability system. *Molecular and Cellular Biology*, 17(1):398–406, 1997.

[108] L. Weill, E. Belloc, F. Bava, and R. Méndez. Translational control by changes in poly(A) tail length: recycling mRNAs. *Nature Structural and Molecular Biology*, 19(6):577–585, 2012.

[109] C. S. Lutz and A. Moreira. Alternative mRNA polyadenylation in eukaryotes: an effective regulator of gene expression. *Wiley Interdisciplinary Reviews: RNA*, 2(1):23–31, 2011.

[110] D. C. Di Giammartino, K. Nishida, and J. L. Manley. Mechanisms and consequences of alternative polyadenylation. *Molecular Cell*, 43(6):853–66, 2011.

[111] A. R. Gruber, G. Martin, W. Keller, and M. Zavolan. Means to an end: mechanisms of alternative polyadenylation of messenger RNA precursors. *Wiley Interdisciplinary Reviews: RNA*, 5(2):183–196, 2014.

[112] D. Zheng and B. Tian. RNA-Binding Proteins in Regulation of Alternative Cleavage and Polyadenylation. *Advances in Experimental Medicine and Biology*, 825:97–127, 2014.

[113] B. Tian and J. L. Manley. Alternative polyadenylation of mRNA precursors. *Nature Reviews Molecular Cell Biology*, 18(1):18–30, 2017.

[114] C. P. Masamha and E. J. Wagner. The contribution of alternative polyadenylation to the cancer phenotype. *Carcinogenesis*, 39(1):2–10, 2018.

[115] Y. Fu, L. Chen, C. Chen, Y. Ge, M. Kang, Z. Song, J. Li, Y. Feng, Z. Huo, G. He, M. Hou, S. Chen, and A. Xu. Crosstalk between alternative polyadenylation and miRNAs in the regulation of protein translational efficiency. *Genome Research*, 28(11):1656–1663, 2018.

[116] N. J. Proudfoot and G. G. Brownlee. 3' Non-coding region sequences in eukaryotic messenger RNA. *Nature*, 263(5574):211–214, 1976.

[117] E. Beaudoing, S. Freier, J. R. Wyatt, J. M. Claverie, and D. Gautheret. Patterns of variant polyadenylation signal usage in human genes. *Genome Research*, 10(7):1001–10, 2000.

[118] N. J. Proudfoot. Ending the message: poly(A) signals then and now. *Genes and Development*, 25(17):1770–1782, 2011.

[119] S. Danckwardt, M. W. Hentze, and A. E. Kulozik. 3' end mRNA processing: molecular mechanisms and implications for health and disease. *The EMBO journal*, 27(3):482–98, 2008.

[120] A. Curinha, S. Oliveira Braz, I. Pereira-Castro, A. Cruz, and A. Moreira. Implications of polyadenylation in health and disease. *Nucleus*, 5(6):508–19, 2014.

[121] K. M. Brown and G. M. Gilmartin. A Mechanism for the Regulation of Pre-mRNA 3' Processing by Human Cleavage Factor Im. *Molecular Cell*, 12(6):1467–1476, 2003.

[122] Q. Yang, M. Coseno, G. M. Gilmartin, and S. Doublié. Crystal Structure of a Human Cleavage Factor CFIm25/CFIm68/RNA Complex Provides an Insight into Poly(A) Site Recognition and RNA Looping. *Structure*, 19(3):368–377, 2011.

[123] S. Millevoi and S. Vagner. Molecular mechanisms of eukaryotic pre-mRNA 3' end processing regulation. *Nucleic Acids Research*, 38(9):2757–74, 2010.

[124] A. Gil and N. J. Proudfoot. A sequence downstream of AAUAAA is required for rabbit $\beta$-globin mRNA 3'-end formation. *Nature*, 312(5993):473–474, 1984.

[125] J. McLauchlan, D. Gaffney, J. L. Whitton, and J. B. Clements. The consensus sequence YGTGTTYY located downstream from the AATAAA signal is required for efficient formation of mRNA 3' termini. *Nucleic Acids Research*, 13(4):1347–68, 1985.

[126] C. C. MacDonald, J. Wilusz, and T. Shenk. The 64-kilodalton subunit of the CstF polyadenylation factor binds to pre-mRNAs downstream of the cleavage site and influences cleavage site location. *Molecular and Cellular Biology*, 14(10):6647–54, 1994.

[127] L. Conway and M. Wickens. A sequence downstream of A-A-U-A-A-A is required for formation of simian virus 40 late mRNA 3' termini in frog oocytes. *Proceedings of the National Academy of Sciences of the United States of America*, 82(12):3949–53, 1985.

[128] D. Rund, C. Dowling, K. Najjar, E. A. Rachmilewitz, H. H. Kazazian, and A. Oppenheim. Two mutations in the beta-globin polyadenylylation signal reveal extended transcripts and new RNA polyadenylylation sites. *Proceedings of the National Academy of Sciences of the United States of America*, 89(10):4324–8, 1992.

[129] Y. Shi, D. C. Di Giammartino, D. Taylor, A. Sarkeshik, W. J. Rice, J. R. Yates, J. Frank, and J. L. Manley. Molecular Architecture of the Human Pre-mRNA 3' Processing Complex. *Molecular Cell*, 33(3):365–376, 2009.

[130] M. Ruepp, D. Schümperli, and S. M. L. Barabino. mRNA 3' end processing and more-multiple functions of mammalian cleavage factor I-68. *Wiley Interdisciplinary Reviews: RNA*, 2(1):79–91, 2011.

[131] S. L. Chan, I. Huppertz, C. Yao, L. Weng, J. J. Moresco, J. R. Yates, J. Ule, J. L. Manley, and Y. Shi. CPSF30 and Wdr33 directly bind to AAUAAA in mammalian mRNA 3' processing. *Genes and Development*, 28(21):2370–80, 2014.

[132] L. Schönemann, U. Kühn, G. Martin, P. Schäfer, A. R. Gruber, W. Keller, M. Zavolan, and E. Wahle. Reconstitution of CPSF active in polyadenylation: Recognition of the polyadenylation signal by WDR33. *Genes and Development*, 28(21):2381–93, 2014.

[133] M. Clerici, M. Faini, R. Aebersold, and M. Jinek. Structural insights into the assembly and polyA signal recognition mechanism of the human CPSF complex. *eLife*, 6:pii:e33111, 2017.

[134] Y. Sun, Y. Zhang, K. Hamilton, J. L. Manley, Y. Shi, T. Walz, and L. Tong. Molecular basis for the recognition of the human AAUAAA polyadenylation signal. *Proceedings of the National Academy of Sciences of the United States of America*, 115(7):E1419–E1428, 2017.

[135] K. Ryan, O. Calvo, and J. L. Manley. Evidence that polyadenylation factor CPSF-73 is the mRNA 3' processing endonuclease. *RNA*, 10(4):565–73, 2004.

[136] C. R. Mandel, S. Kaneko, H. Zhang, D. Gebauer, V. Vethantham, J. L. Manley, and L. Tong. Polyadenylation factor CPSF-73 is the pre-mRNA 3'-end-processing endonuclease. *Nature*, 444(7121):953–6, 2006.

[137] I. Kaufmann, G. Martin, A. Friedlein, H. Langen, and W. Keller. Human Fip1 is a subunit of CPSF that binds to U-rich RNA elements and stimulates poly(A) polymerase. *The EMBO journal*, 23(3):616–26, 2004.

[138] B. Lackford, C. Yao, G. M. Charles, L. Weng, X. Zheng, E.-A. Choi, X. Xie, J. Wan, Y. Xing, J. M. Freudenberg, P. Yang, R. Jothi, G. Hu, and Y. Shi. Fip1 regulates mRNA alternative polyadenylation to promote stem cell self-renewal. *The EMBO Journal*, 33(8):878–889, 2014.

[139] Q. Yang and S. Doublié. Structural biology of poly(A) site definition. *Wiley Interdisciplinary Reviews: RNA*, 2(5):732–47, 2011.

[140] C R Mandel, Y Bai, and L Tong. Protein factors in pre-mRNA 3'-end processing. *Cellular and molecular life sciences : CMLS*, 65(7-8):1099–122, apr 2008.

[141] Y. Takagaki and J. L. Manley. Levels of Polyadenylation Factor CstF-64 Control IgM Heavy Chain mRNA Accumulation and Other Events Associated with B Cell Differentiation. *Molecular Cell*, 2(6):761–771, 1998.

[142] C. Yao, J. Biesinger, J. Wan, L. Weng, Y. Xing, X. Xie, and Y. Shi. Transcriptome-wide analyses of CstF64-RNA interactions in global regulation of mRNA alternative polyadenylation. *Proceedings of the National Academy of Sciences of the United States of America*, 109(46):18773–8, 2012.

[143] G. S. Shankarling, P. W. Coates, B. Dass, and C. C. MacDonald. A family of splice variants of CstF-64 expressed in vertebrate nervous systems. *BMC Molecular Biology*, 10(1):22, 2009.

[144] A. M. Wallace, B. Dass, S. E. Ravnik, V. Tonk, N. A. Jenkins, D. J. Gilbert, N. G. Copeland, and C. C. MacDonald. Two distinct forms of the 64,000 Mr protein of the cleavage stimulation factor are expressed in mouse male germ cells. *Proceedings of the National Academy of Sciences of the United States of America*, 96(12):6763–8, 1999.

[145] B. Dass, K. W. McMahon, N. A. Jenkins, D. J. Gilbert, N. G. Copeland, and C. C. MacDonald. The gene for a variant form of the polyadenylation protein CstF-64 is on chromosome 19 and is expressed in pachytene spermatocytes in mice. *The Journal of Biological Chemistry*, 276(11):8044–50, 2001.

[146] M. Ruepp, C. Schweingruber, N. Kleinschmidt, and D. Schümperli. Interactions of CstF-64, CstF-77, and symplekin: implications on localisation and function. *Molecular Biology of the Cell*, 22(1):91–104, 2011.

[147] T. Kubo, T. Wada, Y. Yamaguchi, A. Shimizu, and H. Handa. Knock-down of 25 kDa subunit of cleavage factor Im in Hela cells alters alternative polyadenylation within 3'-UTRs. *Nucleic Acids Research*, 34(21):6264–71, 2006.

[148] B. L. Sartini, H. Wang, W. Wang, C. F. Millette, and D. L. Kilpatrick. Pre-Messenger RNA Cleavage Factor I (CFIm): Potential Role in Alternative Polyadenylation During Spermatogenesis. *Biology of Reproduction*, 78(3):472–482, 2008.

[149] H. de Vries, U. Rüegsegger, W. Hübner, A. Friedlein, H. Langen, and W. Keller. Human pre-mRNA cleavage factor II(m) contains homologs of yeast proteins and bridges two other cleavage factors. *The EMBO journal*, 19(21):5895–904, 2000.

[150] P. Schäfer, C. Tüting, L. Schönemann, U. Kühn, T. Treiber, N. Treiber, C. Ihling, A. Graber, W. Keller, G. Meister, A. Sinz, and E. Wahle. Reconstitution of mammalian cleavage factor II involved in 3' processing of mRNA precursors. *RNA*, 24(12):1721–1737, 2018.

[151] A. Volanakis, K. Kamieniarz-Gdula, M. Schlackow, and N. J. Proudfoot. WNK1 kinase and the termination factor PCF11 connect nuclear mRNA export with transcription. *Genes and Development*, 31(21):2175–2185, 2017.

[152] K. Kamieniarz-Gdula, M. R. Gdula, K. Panser, N. Brockdorff, A. Pauli, and Proudfoot N. J. Selective Roles of Vertebrate PCF11 in Premature and Full-Length Transcript Termination. *Molecular Cell*, 74(1):158–172.e9., 2019.

[153] Y. Takagaki and J. L. Manley. Complex protein interactions within the human polyadenylation machinery identify a novel component. *Molecular and Cellular Biology*, 20(5):1515–25, 2000.

[154] M. A. Winters and M. Edmonds. A Poly (A) Polymerase from Calf Thymus. *Journal of Biological Chemistry*, 248(13):4756–62, 1973.

[155] U. Kühn, M. Gündel, A. Knoth, Y. Kerwitz, S. Rüdel, and E. Wahle. Poly(A) Tail Length Is Controlled by the Nuclear Poly(A)-binding Protein Regulating the Interaction between Poly(A) Polymerase and the Cleavage and Polyadenylation Specificity Factor. *Journal of Biological Chemistry*, 284(34):22803–22814, 2009.

[156] U. Kühn, J. Buschmann, and E. Wahle. The nuclear poly(A) binding protein of mammals, but not of fission yeast, participates in mRNA polyadenylation. *RNA*, 23(4):473–482, 2017.

[157] L. Weill, E. Belloc, F. A. Bava, and R. Mendez. Translational control by changes in poly(A) tail length: recycling mRNAs. *Nature Structural and Molecular Biology*, 19(6):577–585, 2012.

[158] H. Chang, J. Lim, M. Ha, and V. N. Kim. TAIL-seq: genome-wide determination of poly(A) tail length and 3' end modifications. *Molecular Cell*, 53(6):1044–52, 2014.

[159] N. L. Garneau, J. Wilusz, and C. J. Wilusz. The highways and byways of mRNA decay. *Nature Reviews Molecular Cell Biology*, 8(2):113–126, 2007.

[160] C. R. Eckmann, C. Rammelt, and E. Wahle. Control of poly(A) tail length. *Wiley Interdisciplinary Reviews: RNA*, 2(3):348–361, 2011.

[161] R. K. Singh, D. Liang, U. R. Gajjalaiahvari, M. M. Kabbaj, J. Paik, and A. Gunjan. Excess histone levels mediate cytotoxicity via multiple mechanisms. *Cell Cycle*, 9(20):4236–44, 2010.

[162] D. M. Miles, X. Peñate, T. S. Olmo, F. Jourquin, M. C. Muñoz Centeno, M. Mendoza, M. N. Simon, S. Chavez, and V. Geli. High levels of histones promote whole-genome-duplications and trigger a Swe1WEE1-dependent phosphorylation of Cdc28CDK1. *eLife*, 7:pii: e35337, 2018.

[163] M. Koyama and H. Kurumizaka. Structural diversity of the nucleosome. *The Journal of Biochemistry*, 163(2):85–95, 2017.

[164] D. V. Fyodorov, B. Zhou, A. I. Skoultchi, and Y. Bai. Emerging roles of linker histones in regulating chromatin structure and function. *Nature Reviews Molecular Cell Biology*, 19(3):192–206, 2018.

[165] W. F. Marzluff. Metazoan replication-dependent histone mRNAs: a distinct set of RNA polymerase II transcripts. *Current Opinion in Cell Biology*, 17(3):274–280, 2005.

[166] W. F. Marzluff, E. J. Wagner, and R. J. Duronio. Metabolism and regulation of canonical histone mRNAs: life without a poly(A) tail. *Nature Reviews Genetics*, 9(11):843–854, 2008.

[167] D. Tan, W. F. Marzluff, Z. Dominski, and L. Tong. Structure of Histone mRNA Stem-Loop, Human Stem-Loop Binding Protein, and 3'hExo Ternary Complex. *Science*, 339(6117):318–321, 2013.

[168] Z. Dominski, L. Zheng, R. Sanchez, and W. F. Marzluff. Stem-Loop Binding Protein Facilitates 3'-End Formation by Stabilizing U7 snRNP Binding to Histone Pre-mRNA. *Molecular and Cellular Biology*, 19(5):3561, 1999.

[169] W. F. Marzluff. Metazoan replication-dependent histone mRNAs: a distinct set of RNA polymerase II transcripts. *Current Opinion in Cell Biology*, 17(3):274–280, 2005.

[170] E. C. Scharl and J. A. Steitz. The site of 3' end formation of histone messenger RNA is a fixed distance from the downstream element recognized by the U7 snRNP. *The EMBO journal*, 13(10):2432–40, 1994.

[171] Z. Dominski, J. Sumerel, R. J. Hanson, and W. F. Marzluff. The polyribosomal protein bound to the 3' end of histone mRNA can function in histone pre-mRNA processing. *RNA*, 1(9):915–23, 1995.

[172] R. J. Hanson, J. Sun, D. G. Willis, and W. F. Marzluff. Efficient Extraction and Partial Purification of the Polyribosome-Associated Stem-Loop Binding Protein Bound to the 3' End of Histone mRNA. *Biochemistry*, 35(7):2146–2156, 1996.

[173] Z. F. Wang, M. L. Whitfield, T. C. Ingledue, Z. Dominski, and W. F. Marzluff. The protein that binds the 3' end of histone mRNA: a novel RNA-binding protein required for histone pre-mRNA processing. *Genes and Development*, 10(23):3028–40, 1996.

[174] L. Brooks, S. M. Lyons, J. M. Mahoney, J. D. Welch, Z. Liu, W. F. Marzluff, and M. L. Whitfield. A multiprotein occupancy map of the mRNP on the 3' end of histone mRNAs. *RNA*, 21(11):1943–65, 2015.

[175] J. Zhang, D. Tan, E. F. DeRose, L. Perera, Z. Dominski, W. F. Marzluff, L. Tong, and T. M. T. Hall. Molecular mechanisms for the regulation of histone mRNA stem-loop-binding protein by phosphorylation. *Proceedings of the National Academy of Sciences of the United States of America*, 111(29):E2937–E2946, 2014.

[176] Z. Dominski, J. A. Erkmann, J. A. Greenland, and W. F. Marzluff. Mutations in the RNA binding domain of stem-loop binding protein define separable requirements for RNA binding and for histone pre-mRNA processing. *Molecular and Cellular Biology*, 21(6):2008–17, 2001.

[177] A. Skrajna, X. Yang, K. Bucholc, J. Zhang, T. M. T. Hall, M. Dadlez, W. F. Marzluff, and Z. Dominski. U7 snRNP is recruited to histone pre-mRNA in a FLASH-dependent manner by two separate regions of the stem-loop binding protein. *RNA*, 23(6):938–951, 2017.

[178] K. L. Mowry, J. A. Steitz, Z. Dominski, and L. Tong. Identification of the human U7 snRNP as one of several factors involved in the 3' end maturation of histone premessenger RNAs. *Science*, 238(4834):1682–7, 1987.

[179] K. D. Sullivan, T. E. Mullen, W. F. Marzluff, and E. J. Wagner. Knockdown of SLBP results in nuclear retention of histone mRNA. *RNA*, 15(3):459–72, 2009.

[180] H. Kaygun and W. F. Marzluff. Translation termination is involved in histone mRNA degradation when DNA replication is inhibited. *Molecular and Cellular Biology*, 25(16):6879–6888, 2005.

[181] N. G. Cakmakci, R. S. Lerner, E. J. Wagner, L. Zheng, and W. F. Marzluff. SLIP1, a factor required for activation of histone mRNA translation by the stem-loop binding protein. *Molecular and Cellular Biology*, 28(3):1182–94, 2008.

[182] J. Choe, S. H. Ahn, and Y. K. Kim. The mRNP remodeling mediated by UPF1 promotes rapid degradation of replication-dependent histone mRNA. *Nucleic Acids Research*, 42(14):9334–9349, 2014.

[183] D. Schümperli and R. S. Pillai. The special Sm core structure of the U7 snRNP: far-reaching significance of a small nuclear ribonucleoprotein. *Cellular and Molecular Life Sciences*, 61(19-20):2560–2570, 2004.

[184] Xiao-cui Yang, Brandon D. Burch, Yan Yan, William F. Marzluff, and Zbigniew Dominski. FLASH, a Proapoptotic Protein Involved in Activation of Caspase-8, Is Essential for 3' End Processing of Histone Pre-mRNAs. *Molecular Cell*, 36(2):267–278, oct 2009.

[185] X. Yang, B. Xu, I. Sabath, L. Kunduru, B. D. Burch, W. F. Marzluff, and Z. Dominski. FLASH Is Required for the Endonucleolytic Cleavage of Histone Pre-mRNAs but Is Dispensable for the 5' Exonucleolytic Degradation of the Downstream Cleavage Product. *Molecular and Cellular Biology*, 31(7):1492–1502, 2011.

[186] A. C. Godfrey, A. E. White, D. C. Tatomer, W. F. Marzluff, and R. J. Duronio. The Drosophila U7 snRNP proteins Lsm10 and Lsm11 are required for histone pre-mRNA processing and play an essential role in development. *RNA*, 15(9):1661–72, 2009.

[187] A. Skrajna, X. Yang, M. Dadlez, W. F. Marzluff, and Z. Dominski. Protein composition of catalytically active U7-dependent processing complexes assembled on histone pre-mRNA containing biotin and a photo-cleavable linker. *Nucleic Acids Research*, 46(9):4752–4770, 2018.

[188] K. Anamika, A. Gyenis, L. Poidevin, O. Poch, and L. Tora. RNA polymerase II pausing downstream of core histone genes is different from genes producing polyadenylated transcripts. *PLoS One*, 7(6):e38769, 2012.

[189] O. Gick, A. Kramert, A. Vasserot, and M. L. Birnstiel. Heat-labile regulatory factor is required for 3' processing of histone precursor mRNAs. Technical report, 1987.

[190] N. G. Kolev and J. A. Steitz. Symplekin and multiple other polyadenylation factors participate in 3'-end maturation of histone mRNAs. *Genes and Development*, 19(21):2583–2592, 2005.

[191] V. Romeo, E. Griesbach, and D. Schümperli. CstF64: cell cycle regulation and functional role in 3' end processing of replication-dependent histone mRNAs. *Molecular and Cellular Biology*, 34(23):4272–84, 2014.

[192] N. G. Kolev, T. A. Yario, E. Benson, and J. A. Steitz. Conserved motifs in both CPSF73 and CPSF100 are required to assemble the active endonuclease for histone mRNA 3'-end maturation. *EMBO reports*, 9(10):1013–8, 2008.

[193] I. Sabath, A. Skrajna, X. Yang, M. Dadlez, W. F. Marzluff, and Z. Dominski. 3'-End processing of histone pre-mRNAs in Drosophila: U7 snRNP is associated with FLASH and polyadenylation factors. *RNA*, 19(12):1726–1744, 2013.

[194] I. Pettinati, P. Grzechnik, C. Ribeiro de Almeida, J. Brem, M. A. McDonough, S. Dhir, N. J. Proudfoot, and C. J. Schofield. Biosynthesis of histone messenger RNA employs a specific 3' end endonuclease. *eLife*, 7:pii: e39865, 2018.

[195] W. F. Marzluff, P. Gongidi, K. R. Woods, J. Jin, and L. J. Maltais. The Human and Mouse Replication-Dependent Histone Genes. *Genomics*, 80(5):487–498, 2002.

[196] D. C. Tatomer, E. Terzo, K. P. Curry, H. Salzler, I. Sabath, G. Zapotoczny, D. J. McKay, Z. Dominski, W. F. Marzluff, and R. J. Duronio. Concentrating pre-mRNA processing factors in the histone locus body facilitates efficient histone mRNA biogenesis. *The Journal of Cell Biology*, 213(5):557–570, 2016.

[197] R. J. Duronio and W. F. Marzluff. Coordinating cell cycle-regulated histone gene expression through assembly and function of the Histone Locus Body. *RNA Biology*, 14(6):726–738, 2017.

[198] J. Liu, C. Murphy, M. Buszczak, S. Clatterbuck, R. Goodman, and J. G. Gall. The Drosophila melanogaster Cajal body. *Journal of Cell Biology*, 172(6):875–884, 2006.

[199] A. E. White, B. D. Burch, X. Yang, P. Y. Gasdaska, Z. Dominski, W. F. Marzluff, and R. J. Duronio. Drosophila histone locus bodies form by hierarchical recruitment of components. *The Journal of Cell Biology*, 193(4):677–94, 2011.

[200] D. C. Tatomer, L. F. Rizzardi, K. P. Curry, A. M. Witkowski, W.. F. Marzluff, and R. J. Duronio. Drosophila Symplekin localizes dynamically to the histone locus body and tricellular junctions. *Nucleus*, 5(6):613–625, 2014.

[201] H. R. Salzler, D. C. Tatomer, P. Y. Malek, S. L. McDaniel, A. N. Orlando, W. F. Marzluff, and R. J. Duronio. A sequence in the Drosophila H3-H4 Promoter triggers histone locus body assembly and biosynthesis of replication-coupled histone mRNAs. *Developmental Cell*, 24(6):623–34, 2013.

[202] L. E. Rieder, K. P. Koreski, K. A. Boltz, G. Kuzu, J. A. Urban, S. K. Bowman, A. Zeidman, W. T. Jordan, M. Y. Tolstorukov, W. F. Marzluff, R. J. Duronio, and E. N. Larschan. Histone locus regulation by the Drosophila dosage compensation adaptor protein CLAMP. *Genes and Development*, 31(14):1494–1508, 2017.

[203] X. Ye, Y. Wei, G. Nalepa, and J. W. Harper. The cyclin E/Cdk2 substrate p220(NPAT) is required for S-phase entry, histone gene expression, and Cajal body maintenance in human somatic cells. *Molecular and Cellular Biology*, 23(23):8586–600, 2003.

[204] C. Das, J. K. Tyler, and M. E. A. Churchill. The histone shuffle: histone chaperones in an energetic dance. *Trends in Biochemical Sciences*, 35(9):476–489, 2010.

[205] R. J. Burgess and Z. Zhang. Histone chaperones in nucleosome assembly and human disease. *Nature Structural and Molecular Biology*, 20(1):14–22, 2013.

[206] Z. A. Gurard-Levin, J. Quivy, and G. Almouzni. Histone Chaperones: Assisting Histone Traffic and Nucleosome Dynamics. *Annual Review of Biochemistry*, 83(1):487–517, 2014.

[207] C. M. Hammond, C. B. Strømme, H. Huang, D. J. Patel, and A. Groth. Histone chaperone networks shaping chromatin function. *Nature Reviews Molecular Cell Biology*, 18(3):141–158, 2017.

[208] P. V. Sauer, Y. Gu, W. H. Liu, F. Mattiroli, D. Panne, K. Luger, and M. E. Churchill. Mechanistic insights into histone deposition and nucleosome assembly by the chromatin assembly factor-1. *Nucleic Acids Research*, 46(19):9907–9917, 2018.

[209] M. E. Harris, R. Bohni, M. H. Schneiderman, L. Ramamurthy, D. Schumperli, and W. F. Marzluff. Regulation of histone mRNA in the unperturbed cell cycle: evidence suggesting control at two posttranscriptional steps. *Molecular and Cellular Biology*, 11(5):2416–2424, 1991.

[210] W F Marzluff, E J Wagner, and R J Duronio. Metabolism and regulation of canonical histone mRNAs: life without a poly(A) tail. *Nat Rev Genet*, 9(11):843–854, 2008.

[211] A. Gunjan and A. Verreault. A Rad53 kinase-dependent surveillance mechanism that regulates histone protein levels in S. cerevisiae. *Cell*, 115(5):537–549, 2003.

[212] C. Fletcher, N. Heintz, and R. G. Roeder. Purification and characterization of OTF-1, a transcription factor regulating cell cycle expression of a human histone H2b gene. *Cell*, 51(5):773–781, 1987.

[213] N. K. Kaludov, L. Pabón-Peña, and M. M. Hurt. Identification of a second conserved element within the coding sequence of a mouse H3 histone gene that interacts with nuclear factors and is necessary for normal expression. *Nucleic Acids Research*, 24(3):523–31, 1996.

[214] S. Jaeger, S. Barends, R. Giege, G. Eriani, and F. Martin. Expression of metazoan replication-dependent histone genes. *Biochimie*, 87(9-10):827–834, 2005.

[215] P. Mitra, R. Xie, R. Medina, H. Hovhannisyan, S. K. Zaidi, Y. Wei, J. W. Harper, J. L. Stein, A. J. van Wijnen, and G. S. Stein. Identification of HiNF-P, a key activator of cell cycle-controlled histone H4 genes at the onset of S phase. *Molecular and Cellular Biology*, 23(22):8110–23, 2003.

[216] J. Zhao, B. Dynlacht, T. Imai, T. Hori, and E. Harlow. Expression of NPAT, a novel substrate of cyclin E-CDK2, promotes S-phase entry. *Genes and Development*, 12(4):456–61, 1998.

[217] J. Zhao, B. K. Kennedy, B. D. Lawrence, D. A. Barbie, A. G. Matera, J. A. Fletcher, and E. Harlow. NPAT links cyclin E-Cdk2 to the regulation of replication-dependent histone gene transcription. *Genes and Development*, 14(18):2283–2297, 2000.

[218] T. Ma, B. A. Van Tine, Y. Wei, M. D. Garrett, D. Nelson, P. D. Adams, J. Wang, J. Qin, L. T. Chow, and J. W. Harper. Cell cycle-regulated phosphorylation of p220(NPAT) by cyclin E/Cdk2 in Cajal bodies promotes histone gene transcription. *Genes and Development*, 14(18):2298–313, 2000.

[219] X. Yang, I. Sabath, L. Kunduru, A. J. van Wijnen, W. F. Marzluff, and Z. Dominski. A Conserved Interaction That Is Essential for the Biogenesis of Histone Locus Bodies. *Journal of Biological Chemistry*, 289(49):33767–33782, 2014.

[220] W. S. Aik, M. H. Lin, D. Tan, A. Tripathy, W. F. Marzluff, Z. Dominski, C. Y. Chou, and L. Tong. The N-terminal domains of FLASH and Lsm11 form a 2:1 heterotrimer for histone pre-mRNA 3'-end processing. *PLoS One*, 12(10):e0186034, 2017.

[221] M. L. Whitfield, L. X. Zheng, A. Baldwin, T. Ohta, M. M. Hurt, and W. F. Marzluff. Stem-loop binding protein, the protein that binds the 3' end of histone mRNA, is cell cycle regulated by both translational and posttranslational mechanisms. *Molecular and Cellular Biology*, 20(12):4188–4198, 2000.

[222] L. Zheng, Z. Dominski, X. Yang, P. Elms, C. S. Raska, C. H. Borchers, and W. F. Marzluff. Phosphorylation of stem-loop binding protein (SLBP) on two threonines triggers degradation of SLBP, the sole cell cycle-regulated factor required for regulation of histone mRNA processing, at the end of S phase. *Molecular and Cellular Biology*, 23(5):1590–1601, 2003.

[223] J. F. Dankert, G. Rona, L. Clijsters, P. Geter, J. R. Skaar, K. Bermudez-Hernandez, E. Sassani, D. Fenyo, B. Ueberheide, R. Schneider, and M. Pagano. Cyclin F-Mediated Degradation of SLBP Limits H2A.X Accumulation and Apoptosis upon Genotoxic Stress in G2. *Molecular Cell*, 64(3):507–519, 2016.

[224] R. Sànchez and W. F. Marzluff. The stem-loop binding protein is required for efficient translation of histone mRNA in vivo and in vitro. *Molecular and Cellular Biology*, 22(20):7093–104, 2002.

[225] S. A. Meaux, C. E. Holmquist, and W. F. Marzluff. Role of oligouridylation in normal metabolism and regulated degradation of mammalian histone mRNAs. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 373(1762):20180170, 2018.

[226] R. Bandyopadhyay, G. Stein, and J. Stein. Coordinate turnover of nuclear and cytoplasmic histone messenger RNA following inhibition of DNA replication in HeLa S3 cells. *Biochemistry*, 26(10):2938–44, 1987.

[227] R. A. Graves, N. B. Pandey, N. Chodchoy, and W. F. Marzluff. Translation is required for regulation of histone mRNA degradation. *Cell*, 48(4):615–26, 1987.

[228] H. Kaygun and W. F. Marzluff. Regulated degradation of replication-dependent histone mRNAs requires both ATR and Upf1. *Nature Structural and Molecular Biology Molecular Biology*, 12(9):794–800, sep 2005.

[229] P. E. Lackey, J. D. Welch, and W. F. Marzluff. TUT7 catalyzes the uridylation of the 3' end for rapid degradation of histone mRNA. *RNA*, 22(11):1673–1688, 2016.

[230] Sundaresan Tharun. Lsm1-7-Pat1 complex: a link between 3' and 5'-ends in mRNA decay? *RNA biology*, 6(3):228–32, 2009.

[231] T E Mullen and W F Marzluff. Degradation of histone mRNA requires oligouridylation followed by decapping and simultaneous degradation of the mRNA both 5' to 3' and 3' to 5'. *Genes Dev*, 22(1):50–65, 2008.

[232] J. A. Erkmann, R. Sànchez, N. Treichel, W. F. Marzluff, and U. Kutay. Nuclear export of metazoan replication-dependent histone mRNAs is dependent on RNA length and is mediated by TAP. *RNA*, 11(1):45–58, 2005.

[233] H. von Moeller, R. Lerner, A. Ricciardi, C. Basquin, W. F. Marzluff, and E. Conti. Structural and biochemical studies of SLIP1–SLBP identify DBP5 and eIF3g as SLIP1-binding proteins. *Nucleic Acids Research*, 41(16):7960–7971, 2013.

[234] J. Choe, K. M. Kim, S. Park, Y. K. Lee, O. Song, M. K. Kim, B. Lee, H. K. Song, and Y. K. Kim. Rapid degradation of replication-dependent histone mRNAs largely occurs on mRNAs bound by nuclear cap-binding proteins 80 and 20. *Nucleic Acids Research*, 41(2):1307–1318, 2013.

[235] S. Henikoff and K. Ahmad. Assembly of Variant Histones into Chromatin. *Annual Review of Cell and Developmental Biology*, 21:133–53, 2005.

[236] C. R. Clapier and B. R. Cairns. The Biology of Chromatin Remodeling Complexes. *Annual Review of Biochemistry*, 78:273–304, 2009.

[237] C. R. Clapier, J. Iwasa, B. R. Cairns, and C. L. Peterson. Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes. *Nature Reviews Molecular Cell Biology*, 18(7):407–422, 2017.

[238] M. Sancho, E. Diani, M. Beato, and A. Jordan. Depletion of Human Histone H1 Variants Uncovers Specific Roles in Gene Expression and Cell Growth. *PLoS Genetics*, 4(10):e1000227, 2008.

[239] S. B. Hake and C. David Allis. Histone H3 variants and their potential role in indexing mammalian genomes: The "H3 barcode hypothesis". *Proceedings of the National Academy of Sciences of the United States of America*, 103(17):6428–35, 2006.

[240] P. B. Talbert and S. Henikoff. Histone variants - ancient wrap artists of the epigenome. *Nature Reviews Molecular Cell Biology*, 11(4):264–275, 2010.

[241] S. Mendiratta, A. Gatto, and G. Almouzni. Histone supply: Multitiered regulation ensures chromatin dynamics throughout the cell cycle. *The Journal of cell biology*, 218(1):39–54, 2019.

[242] E. Sullivan, C. Santiago, E. D. Parker, Z. Dominski, X. Yang, D. J. Lanzotti, T. C. Ingledue, W. F. Marzluff, and R. J. Duronio. Drosophila stem loop binding protein coordinates accumulation of mature histone mRNA with cell cycle progression. *Genes and Development*, 15(2):173–87, 2001.

[243] A. C. Godfrey, J. M. Kupsco, B. D. Burch, R. M. Zimmerman, Z. Dominski, W. F. Marzluff, and R. J. Duronio. U7 snRNA mutations in Drosophila block histone pre-mRNA processing and disrupt oogenesis. *RNA*, 12(3):396–409, 2006.

[244] J. Hsin, A. Sheth, and J. L. Manley. RNAP II CTD Phosphorylated on Threonine-4 Is Required for Histone mRNA 3' End Processing. *Science*, 334(6056):683–686, 2011.

[245] J. J. Gruber, S. H. Olejniczak, J. Yong, G. La Rocca, G. Dreyfuss, and C. B. Thompson. Ars2 Promotes Proper Replication-Dependent Histone mRNA 3' End Formation. *Molecular Cell*, 45(1):87–98, 2012.

[246] J. Pirngruber, A. Shchebet, L. Schreiber, E. Shema, N. Minsky, R. D. Chapman, D. Eick, Y. Aylon, M. Oren, and S. A. Johnsen. CDK9 directs H2B monoubiquitination and controls replication-dependent histone mRNA 3'-end processing. *EMBO Reports*, 10(8):894–900, 2009.

[247] L. J. Farber, E. J. Kort, P. Wang, J. Chen, and B. T. Teh. The tumor suppressor parafibromin is required for posttranscriptional processing of histone mRNA. *Molecular Carcinogenesis*, 49(3):215–23, 2009.

[248] N. Chodchoy, N. B. Pandey, and W. F. Marzluff. An intact histone 3'-processing site is required for transcription termination in a mouse histone H2a gene. *Molecular and Cellular Biology*, 11(1):497–509, 1991.

[249] D. J. Lanzotti, H. Kaygun, X. Yang, R. J. Duronio, and W. F. Marzluff. Developmental control of histone mRNA and dSLBP synthesis during Drosophila embryogenesis and the role of dSLBP in histone mRNA 3' end processing in vivo. *Molecular and Cellular Biology*, 22(7):2267–2282, 2002.

[250] V. Kari, O. Karpiuk, B. Tieg, M. Kriegs, E. Dikomey, H. Krebber, Y. Begus-Nahrmann, and S. A. Johnsen. A Subset of Histone H2B Genes Produces Polyadenylated mRNAs under a Variety of Cellular Conditions. *PLoS ONE*, 8(5):e63745, 2013.

[251] S. M. Lyons, C. H. Cunningham, J. D. Welch, B. Groh, A. Y. Guo, B. Wei, M. L. Whitfield, Y. Xiong, and W. F. Marzluff. A subset of replication-dependent histone mRNAs are expressed as polyadenylated RNAs in terminally differentiated tissues. *Nucleic Acids Research*, 44(19):9190–9205, 2016.

[252] C. Mannironi, W. M. Bonner, and C. L. Hatch. H2A.X. a histone isoprotein with a conserved C-terminal sequence, is encoded by a novel mRNA with both DNA replication type and polyA 3' processing signals. *Nucleic Acids Research*, 17(22):9113–9126, 1989.

[253] V. S. Ivanova, D. Zimonjic, N. Popescu, and W. M. Bonner. Chromosomal localization of the human histone H2A.X gene to 11q23.2-q23.3 by fluorescence in situ hybridization. *Human Genetics*, 94(3):303–306, 1994.

[254] D. M. S. Pinto and A. Flaus. Structure and Function of Histone H2AX. *Subcellular Biochemistry*, 50:55–78, 2010.

[255] V S Ivanova, C L Hatch, and W M Bonner. Characterization of the human histone H2A.X gene. Comparison of its promoter with other H2A gene promoters. *J Biol Chem*, 269(39):24189–24194, 1994.

[256] W. M. Bonner, C. Mannironi, A. Orr, D. R. Pilch, and C. L. Hatch. Histone H2A.X gene transcription is regulated differently than transcription of other replication-linked histone genes. *Molecular and Cellular Biology*, 13(2):984–992, 1993.

[257] L. L. Baumbach, F. Marashi, M. Plumb, G. Stein, and J. Stein. Inhibition of DNA replication coordinately reduces cellular levels of core and H1 histone mRNAs: requirement for protein synthesis. *Biochemistry*, 23(8):1618–25, 1984.

[258] H. S. Malik and S. Henikoff. Phylogenomics of the nucleosome. *Nature Structural and Molecular Biology*, 10(11):882–891, 2003.

[259] E. P. Rogakou, D. R. Pilch, A. H. Orr, V. S. Ivanova, and W. M. Bonner. DNA double-stranded breaks induce histone H2AX phosphorylation on serine 139. *Journal of Biological Chemistry*, 273(10):5858–5868, 1998.

[260] A. A. Hatimy, M. J. G. Browne, A. Flaus, and S. M. M. Sweet. Histone H2AX Y142 phosphorylation is a low abundance modification. *International Journal of Mass Spectrometry*, 391:139–145, 2015.

[261] S. Burma, B. P. Chen, M. Murphy, A. Kurimasa, and D. J. Chen. ATM Phosphorylates Histone H2AX in Response to DNA Double-strand Breaks. *Journal of Biological Chemistry*, 276(45):42462–7, 2001.

[262] I. M. Ward and J. Chen. Histone H2AX Is Phosphorylated in an ATR-dependent Manner in Response to Replicational Stress. *Journal of Biological Chemistry*, 276(51):47759–62, 2001.

[263] T. Furuta, H. Takemura, Z. Y. Liao, G. J. Aune, C. Redon, O. A. Sedelnikova, D. R. Pilch, E. P. Rogakou, A. Celeste, H. T. Chen, A. Nussenzweig, M. I. Aladjem, W. M. Bonner, and Y. Pommier. Phosphorylation of histone H2AX and activation of Mre11, Rad50, and Nbs1 in response to replication-dependent DNA double-strand breaks induced by mammalian DNA topoisomerase I cleavage complexes. *Journal of Biological Chemistry*, 278(22):20303–20312, 2003.

[264] T. Stiff, M. O'Driscoll, N. Rief, K. Iwabuchi, M. Löbrich, and P. A. Jeggo. ATM and DNA-PK Function Redundantly to Phosphorylate H2AX after Exposure to Ionizing Radiation. *Cancer Research*, 64(7):2390–6, 2004.

[265] M. Stucki, J. A. Clapperton, D. Mohammad, M. B. Yaffe, S. J. Smerdon, and S. P. Jackson. MDC1 directly binds phosphorylated histone H2AX to regulate cellular responses to DNA double-strand breaks. *Cell*, 123(7):1213–26, 2005.

[266] C. Lukas, F. Melander, M. Stucki, J. Falck, S. Bekker-Jensen, M. Goldberg, Y. Lerenthal, S. P. Jackson, J. Bartek, and J. Lukas. Mdc1 couples DNA double-strand break recognition by Nbs1 with its H2AX-dependent chromatin retention. *The EMBO Journal*, 23:2674–2683, 2004.

[267] T. Uziel, Y. Lerenthal, L. Moyal, Y. Andegeko, L. Mittelman, and Y. Shiloh. Requirement of the MRN complex for ATM activation by DNA damage. *The EMBO Journal*, 22(20):5612–21, 2003.

[268] E. P. Rogakou, C. Boon, C. Redon, and W. M. Bonner. Megabase chromatin domains involved in DNA double-strand breaks in vivo. *Journal of Cell Biology*, 146(5):905–916, 1999.

[269] O. A. Sedelnikova, E. P. Rogakou, I. G. Panyutin, and W. M. Bonner. Quantitative detection of (125)IdU-induced DNA double-strand breaks with gamma-H2AX antibody. *Radiation Research*, 158(4):486–492, 2002.

[270] T. T. Paull, E. P. Rogakou, V. Yamazaki, C. U. Kirchgessner, M. Gellert, and W. M. Bonner. A critical role for histone H2AX in recruitment of repair factors to nuclear foci after DNA damage. *Current Biology*, 10(15):886–95, 2000.

[271] A. Celeste, S. Petersen, P. J. Romanienko, O. Fernandez-Capetillo, H. T. Chen, O. A. Sedelnikova, B. Reina-San-Martin, V. Coppola, Eric Meffre, M. J. Difilippantonio, C. Redon, D. R. Pilch, A. Olaru, M. Eckhaus, R. D. Camerini-Otero, L. Tessarollo, F. Livak, K. Manova, W. M. Bonner, M. C. Nussenzweig, and A. Nussenzweig. Genomic instability in mice lacking histone H2AX. *Science*, 296(5569):922–7, 2002.

[272] A. Celeste, O. Fernandez-Capetillo, M. J. Kruhlak, D. R. Pilch, D. W. Staudt, A. Lee, R. F. Bonner, W. M. Bonner, and A. Nussenzweig. Histone H2AX phosphorylation is dispensable for the initial recognition of DNA breaks. *Nature Cell Biology*, 5(7):675–679, 2003.

[273] M. J. Kruhlak, A. Celeste, G. Dellaire, O. Fernandez-Capetillo, W. G. Müller, J. G. McNally, D. P. Bazett-Jones, and A. Nussenzweig. Changes in chromatin structure and mobility in living cells at sites of DNA double-strand breaks. *The Journal of Cell Biology*, 172(6):823–34, 2006.

[274] O. Fernandez-Capetillo, A. Lee, M. Nussenzweig, and A. Nussenzweig. H2AX: The histone guardian of the genome. *DNA Repair*, 3(8-9):959–67, 2004.

[275] A. J. Nakamura, V. A. Rao, Y. Pommier, and W. M. Bonner. The complexity of phosphorylated H2AX foci formation and DNA repair assembly at DNA double-strand breaks. *Cell Cycle*, 9(2):389–397, 2010.

[276] D. Chowdhury, M. Keogh, H. Ishii, C. L. Peterson, S. Buratowski, and J. Lieberman. Gamma-H2AX Dephosphorylation by Protein Phosphatase 2A Facilitates DNA Double-Strand Break Repair. *Molecular Cell*, 20(5):801–809, 2005.

[277] M. Keogh, J. Kim, M. Downey, J. Fillingham, D. Chowdhury, J. C. Harrison, M. Onishi, N. Datta, S. Galicia, A. Emili, J. Lieberman, X. Shen, S. Buratowski, J. E. Haber, D. Durocher, J. F. Greenblatt, and N. J. Krogan. A phosphatase complex that dephosphorylates $\gamma$H2AX regulates DNA damage checkpoint recovery. *Nature*, 439(7075):497–501, 2006.

[278] A. Nussenzweig and T. Paull. Tails of histones lost. *Nature*, 439(7075):406–407, 2006.

[279] W. M. Bonner, C. E. Redon, J. S. Dickey, A. J. Nakamura, O. A. Sedelnikova, S. Solier, and Y. Pommier. GammaH2AX and cancer. *Nature Reviews Cancer*, 8(12):957–967, 2008.

[280] Z. Li, J. Yang, and H. Huang. Oxidative stress induces H2AX phosphorylation in human spermatozoa. *FEBS Letters*, 580(26):6161–6168, 2006.

[281] E. P. Rogakou, W. Nieves-Neira, C. Boon, Y. Pommier, and W. M. Bonner. Initiation of DNA fragmentation during apoptosis induces phosphorylation of H2AX histone at serine 139. *The Journal of Biological Chemistry*, 275(13):9390–5, 2000.

[282] M. K. Zeman and K. A. Cimprich. Causes and consequences of replication stress. *Nature Cell Biology*, 16(1):2–9, 2014.

[283] S. Petersen, R. Casellas, B. Reina-San-Martin, H. T. Chen, M. J. Difilippantonio, P. C. Wilson, L. Hanitsch, A. Celeste, M. Muramatsu, D. R. Pilch, C. Redon, T. Ried, W. M. Bonner, T. Honjo, M. C. Nussenzweig, and A. Nussenzweig. AID is required to initiate Nbs1/gamma-H2AX focus formation and mutations at sites of class switching. *Nature*, 414(6864):660–665, 2001.

[284] P. L. Olive. Endogenous DNA breaks: gammaH2AX and the role of telomeres. *Aging*, 1(2):154–6, 2009.

[285] M. Wakasugi, T. Sasaki, M. Matsumoto, M. Nagaoka, K. Inoue, M. Inobe, K. Horibata, K. Tanaka, and T. Matsunaga. Nucleotide Excision Repair-dependent DNA Double-strand Break Formation and ATM Signaling Activation in Mammalian Quiescent Cells. *Journal of Biological Chemistry*, 289(41):28730–28737, 2014.

[286] H. T. Chen, A. Bhandoola, M. J. Difilippantonio, J. Zhu, M. J. Brown, X. Tai, E. P. Rogakou, T. M. Brotz, W. M. Bonner, T. Ried, and A. Nussenzweig. Response to RAG-mediated V(D)J cleavage by NBS1 and gamma-H2AX. *Science*, 290(5498):1962–5, 2000.

[287] C. H. Bassing, K. F. Chua, J. Sekiguchi, H. Suh, S. R. Whitlow, J. C. Fleming, B. C. Monroe, D. N. Ciccone, C. Yan, K. Vlasakova, D. M. Livingston, D. O. Ferguson, R. Scully, and F. W. Alt. Increased ionizing radiation sensitivity and genomic instability in the absence of histone H2AX. *Proceedings of the National Academy of Sciences of the United States of America*, 99(12):8173–8178, 2002.

[288] R. Daniel, J. Ramcharan, E. Rogakou, K. D. Taganov, J. G. Greger, W. Bonner, A. Nussenzweig, R. A. Katz, and A. M. Skalka. Histone H2AX is phosphorylated at sites of retroviral DNA integration but is dispensable for postintegration repair. *Journal of Biological Chemistry*, 279(44):45810–45814, 2004.

[289] T. Stiff, S. A. Walker, K. Cerosaletti, A. A. Goodarzi, E. Petermann, P. Concannon, M. O'Driscoll, and P. A. Jeggo. ATR-dependent phosphorylation and activation of ATM in response to UV treatment or replication fork stalling. *The EMBO Journal*, 25(24):5775–5782, 2006.

[290] E. Moeglin, D. Desplancq, S. Conic, M. Oulad-Abdelghani, A. Stoessel, M. Chiper, M. Vigneron, P. Didier, L. Tora, and E. Weiss. Uniform Widespread Nuclear Phosphorylation of Histone H2AX Is an Indicator of Lethal DNA Replication Stress. *Cancers*, 11(3):355, 2019.

[291] C. H. Bassing, H. Suh, D. O. Ferguson, K. F. Chua, J. Manis, M. Eckersdorff, M. Gleason, R. Bronson, C. Lee, and F. W. Alt. Histone H2AX: A Dosage-Dependent Suppressor of Oncogenic Translocations and Tumors. *Cell*, 114(3):359–370, 2003.

[292] M. J. Thirman, H. J. Gill, R. C. Burnett, D. Mbangkollo, N. R. McCabe, H. Kobayashi, S. Ziemin-van der Poel, Y. Kaneko, R. Morgan, A. A. Sandberg, R. S. K. Chaganti, R. A. Larson, M. M. Le Beau, M. O. Diaz, and J. D. Rowley. Rearrangement of the MLL Gene in Acute Lymphoblastic and Acute Myeloid Leukemias with 11q23 Chromosomal Translocations. *New England Journal of Medicine*, 329(13):909–914, 1993.

[293] P. J. Cook, B. G. Ju, F. Telese, X. Wang, C. K. Glass, and M. G. Rosenfeld. Tyrosine dephosphorylation of H2AX modulates apoptosis and survival decisions. *Nature*, 458(7238):591–596, 2009.

[294] A. Xie, S. Odate, G. Chandramouly, and R. Scully. H2AX post-translational modifications in the ionizing radiation response and homologous recombination. *Cell Cycle*, 9(17):3602–10, 2010.

[295] J. B. Pinder, K. M. Attwood, and G. Dellaire. Reading, writing, and repair: the role of ubiquitin and the ubiquitin-like proteins in DNA damage signaling and repair. *Frontiers in Genetics*, 4:45, 2013.

[296] M. W. Luczak and A. Zhitkovich. Monoubiquitinated gamma-H2AX: Abundant product and specific biomarker for non-apoptotic DNA double-strand breaks. *Toxicology and Applied Pharmacology*, 355:238–246, 2018.

[297] V. Turinetto and C. Giachino. Multiple facets of histone variant H2AX: a DNA double-strand-break marker with several biological functions. *Nucleic Acids Research*, 43(5):2489–2498, 2015.

[298] U. Weyemi, C. E. Redon, R. Choudhuri, T. Aziz, D. Maeda, M. Boufraqech, P. R. Parekh, T. K. Sethi, M. Kasoji, N. Abrams, A. Merchant, V. N. Rajapakse, and W. M. Bonner. The histone variant H2A.X is a regulator of the epithelial-mesenchymal transition. *Nature Communications*, 7:10711, 2016.

[299] U. Weyemi, C. E. Redon, T. K. Sethi, A. S. Burrell, P. Jailwala, M. Kasoji, N. Abrams, A. Merchant, and W. M. Bonner. Twist1 and Slug mediate H2AX-regulated epithelial-mesenchymal transition in breast cells. *Cell Cycle*, 15(18):2398–2404, 2016.

[300] O Fernandez-Capetillo, S K Mahadevaiah, A Celeste, P J Romanienko, R D Camerini-Otero, W M Bonner, K Manova, P Burgoyne, and A Nussenzweig. H2AX is required for chromatin remodeling and inactivation of sex chromosomes in male mouse meiosis. *Dev Cell*, 4(4):497–508, 2003.

[301] B. P. Chadwick and T. F. Lane. BRCA1 associates with the inactive X chromosome in late S-phase, coupled with transient H2AX phosphorylation. *Chromosoma*, 114(6):432–439, 2005.

[302] Y. Ichijima, R. Sakasai, N. Okita, K. Asahina, S. Mizutani, and H. Teraoka. Phosphorylation of histone H2AX at M phase in human cells without DNA damage response. *Biochemical and Biophysical Research Communications*, 336(3):807–812, 2005.

[303] Y. Eliezer, L. Argaman, M. Kornowski, M. Roniger, and M. Goldberg. Interplay between the DNA damage proteins MDC1 and ATM in the regulation of the spindle assembly checkpoint. *The Journal of Biological Chemistry*, 289(12):8182–93, 2014.

[304] G. W. Charville and T. A. Rando. The mortal strand hypothesis: non-random chromosome inheritance and the biased segregation of damaged DNA. *Seminars in Cell and Developmental Biology*, 24(8-9):653–60, 2013.

[305] F. Dotiwala, J. C. Harrison, S. Jain, N. Sugawara, and J. E. Haber. Mad2 Prolongs DNA Damage Checkpoint Arrest Caused by a Double-Strand Break via a Centromere-Dependent Mechanism. *Current Biology*, 20(4):328–332, 2010.

[306] G. R. Kafer, S. A. Lehnert, M. Pantaleon, P. L. Kaye, and R. J. Moser. Expression of genes coding for histone variants and histone-associated proteins in pluripotent stem cells and mouse preimplantation embryos. *Gene Expression Patterns*, 10(6):299–305, 2010.

[307] V. Turinetto, L. Orlando, Y. Sanchez-Ripoll, B. Kumpfmueller, M. P. Storm, P. Porcedda, V. Minieri, S. Saviozzi, L. Accomasso, E. Cibrario Rocchietti, K. Moorwood, P. Circosta, A. Cignetti, M. J. Welham, and C. Giachino. High Basal gamma-H2AX Levels Sustain Self-Renewal of Mouse Embryonic and Induced Pluripotent Stem Cells. *Stem Cells*, 30(7):1414–1423, 2012.

[308] T. Wu, Y. Liu, D. Wen, Z. Tseng, M. Tahmasian, M. Zhong, S. Rafii, M. Stadtfeld, K. Hochedlinger, and A. Xiao. Histone Variant H2A.X Deposition Pattern Serves as a Functional Epigenetic Mark for Distinguishing the Developmental Potentials of iPSCs. *Cell Stem Cell*, 15(3):281–294, 2014.

[309] F. Rodier, J. Coppé, C. K. Patil, W. A. M. Hoeijmakers, D. P. Muñoz, S. R. Raza, A. Freund, E. Campeau, A. R. Davalos, and J. Campisi. Persistent DNA damage signalling triggers senescence-associated inflammatory cytokine secretion. *Nature Cell Biology*, 11(8):973–979, 2009.

[310] F. Rodier, D. P. Muñoz, R. Teachenor, V. Chu, O. Le, D. Bhaumik, J. Coppé, E. Campeau, C. M. Beauséjour, S. Kim, A. R. Davalos, and J. Campisi. DNA-SCARS: distinct nuclear structures that sustain damage-induced senescence growth arrest and inflammatory cytokine secretion. *Journal of Cell Science*, 124(1):68–81, 2011.

[311] A. Lal, Y. Pan, F. Navarro, D. M. Dykxhoorn, L. Moreau, E. Meire, Z. Bentwich, J. Lieberman, and D. Chowdhury. miR-24-mediated downregulation of H2AX suppresses DNA repair in terminally differentiated blood cells. *Nature Structural and Molecular Biology*, 16(5):492–498, 2009.

[312] Y. Wang, J. W. Huang, M. Li, W. K. Cavenee, P. S. Mitchell, X. Zhou, M. Tewari, F. B. Furnari, and T. Taniguchi. MicroRNA-138 modulates DNA damage response by repressing histone H2AX expression. *Molecular Cancer Research*, 9(8):1100–1111, 2011.

[313] H. Yang, J. Luo, Z. Liu, R. Zhou, and H. Luo. MicroRNA-138 Regulates DNA Damage Response in Small Cell Lung Cancer Cells by Directly Targeting H2AX. *Cancer Investigation*, 33(4):126–136, 2015.

[314] M. J. De Brabander, R. M. Van de Veire, F. E. Aerts, M. Borgers, and P. A. Janssen. The effects of methyl (5-(2-thienylcarbonyl)-1H-benzimidazol-2-yl) carbamate, (R 17934; NSC 238159), a new synthetic antitumoral drug interfering with microtubules, on mammalian cells cultured in vitro. *Cancer Research*, 36(3):905–916, 1976.

[315] G. Pedrali-Noy and S. Spadari. Effect of aphidicolin on viral and human DNA polymerases. *Biochemical and Biophysical Research Communications*, 88(4):1194–1202, 1979.

[316] A. G. Baranovskiy, N. D. Babayeva, Y. Suwa, J. Gu, Y. I. Pavlov, and T. H. Tahirov. Structural basis for inhibition of DNA replication by aphidicolin. *Nucleic Acids Research*, 42(22):14013–21, 2014.

[317] N. Xeros. Deoxyriboside control and synchronization of mitosis. *Nature*, 194:682–683, 1962.

[318] H. Liu and J. H. Naismith. An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. *BMC biotechnology*, 8(91), 2008.

[319] J. A. Gagnon, E. Valen, S. B. Thyme, P. Huang, L. Ahkmetova, A. Pauli, T. G. Montague, S. Zimmerman, C. Richter, and A. F. Schier. Efficient Mutagenesis by Cas9 Protein-Mediated Oligonucleotide Insertion and Large-Scale Assessment of Single-Guide RNAs. *PLoS ONE*, 9(5):e98186, 2014.

[320] A. Montecucco, F. Zanetta, and G. Biamonti. Molecular mechanisms of etoposide. *EXCLI Journal*, 14:95–108, 2015.

[321] A. Muslimović, S. Nyström, Y. Gao, and O. Hammarsten. Numerical analysis of etoposide induced DNA breaks. *PLoS One*, 4(6):e5859, 2009.

[322] Y. Atsumi, Y. Minakawa, M. Ono, S. Dobashi, K. Shinohe, A. Shinohara, S. Takeda, M. Takagi, N. Takamatsu, H. Nakagama, H. Teraoka, and K. Yoshioka. ATM and SIRT6/SNF2H Mediate Transient H2AX Stabilization When DSBs Form by Blocking HUWE1 to Allow Efficient gammaH2AX Foci Formation. *Cell Reports*, 13(12):2728–2740, 2015.

[323] H. Kaygun and W. F. Marzluff. Translation termination is involved in histone mRNA degradation when DNA replication is inhibited. *Molecular and Cellular Biology*, 25(16):6879–88, 2005.

[324] U. M. Bond, T. A. Yario, and J. A. Steitz. Multiple processing-defective mutations in a mammalian histone pre-mRNA are suppressed by compensatory changes in U7 RNA both in vivo and in vitro. *Genes and Development*, 5(9):1709–1722, 1991.

[325] Yu-Shan Chen and Xiao-Bo Qiu. Transcription-Coupled Replacement of Histones: Degradation or Recycling? *Journal of Genetics and Genomics*, 39(11):575–580, nov 2012.

[326] V Kari, O Karpiuk, B Tieg, M Kriegs, E Dikomey, H Krebber, Y Begus-Nahrmann, and S A Johnsen. A subset of histone H2B genes produces polyadenylated mRNAs under a variety of cellular conditions. *PLoS One*, 8(5):e63745, 2013.

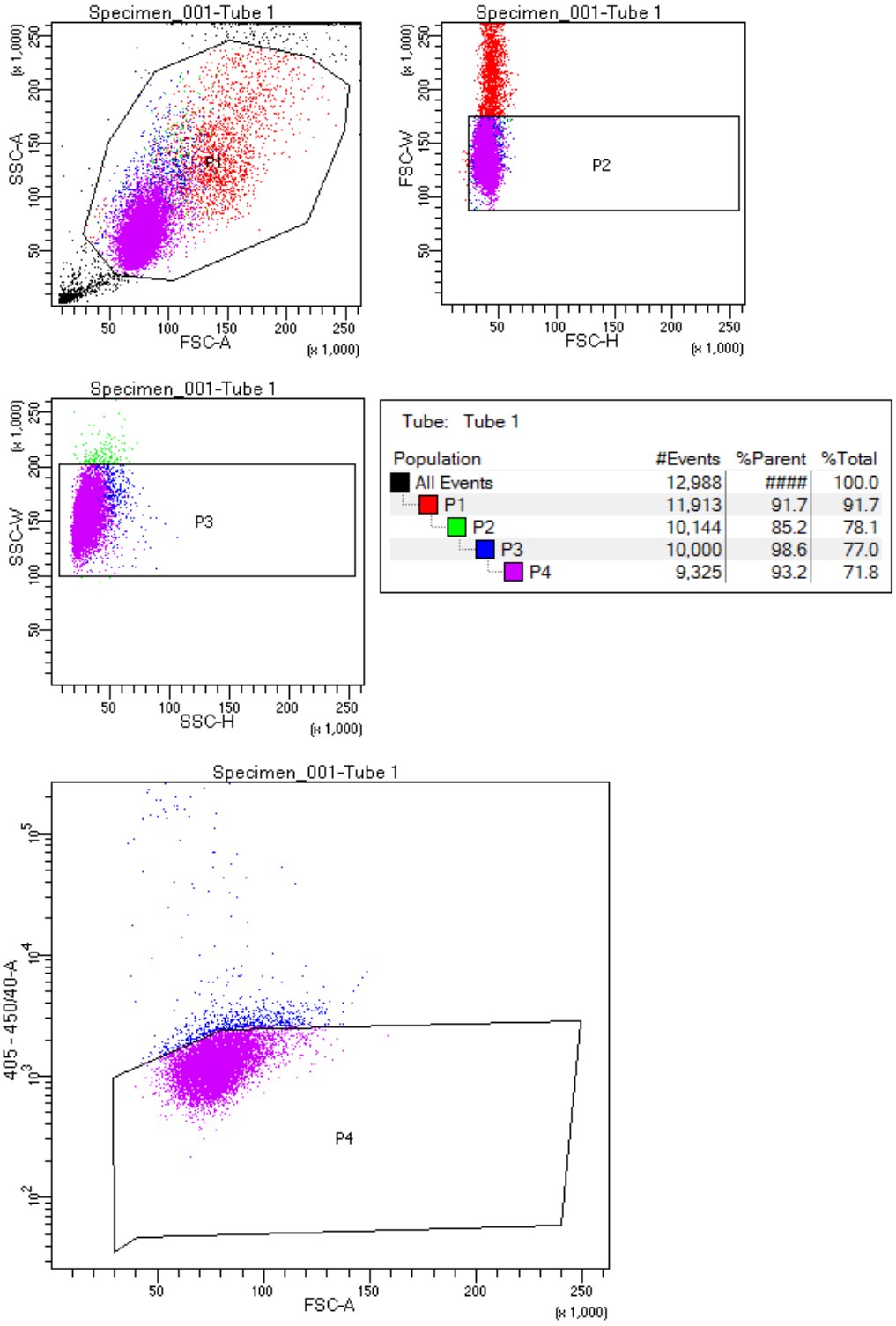
[327] D. Dominguez, Y. H. Tsai, N. Gomez, D. K. Jha, I. Davis, and Z. Wang. A high-resolution transcriptome map of cell cycle reveals novel connections between periodic genes and cancer. *Cell Research*, 26(8):946–962, 2016.

[328] Y. Atsumi, H. Fujimori, H. Fukuda, A. Inase, K. Shinohe, Y. Yoshioka, M. Shikanai, Y. Ichijima, J. Unno, S. Mizutani, N. Tsuchiya, Y. Hippo, H. Nakagama, M. Masutani, H. Teraoka, and K. Yoshioka. Onset of Quiescence Following p53 Mediated Down-Regulation of H2AX in Normal Cells. *PLoS One*, 6(8):e23432, 2011.

[329] J. Seo, S. C. Kim, H. S. Lee, J. K. Kim, H. J. Shon, N. L. Salleh, K. V. Desai, J. H. Lee, E. S. Kang, J. S. Kim, and J. K. Choi. Genome-wide profiles of H2AX and gamma-H2AX differentiate endogenous and exogenous DNA damage hotspots in human cells. *Nucleic Acids Research*, 40(13):5965–5974, 2012.

# Supplementary Figures

BD FACSDiva 8.0.2



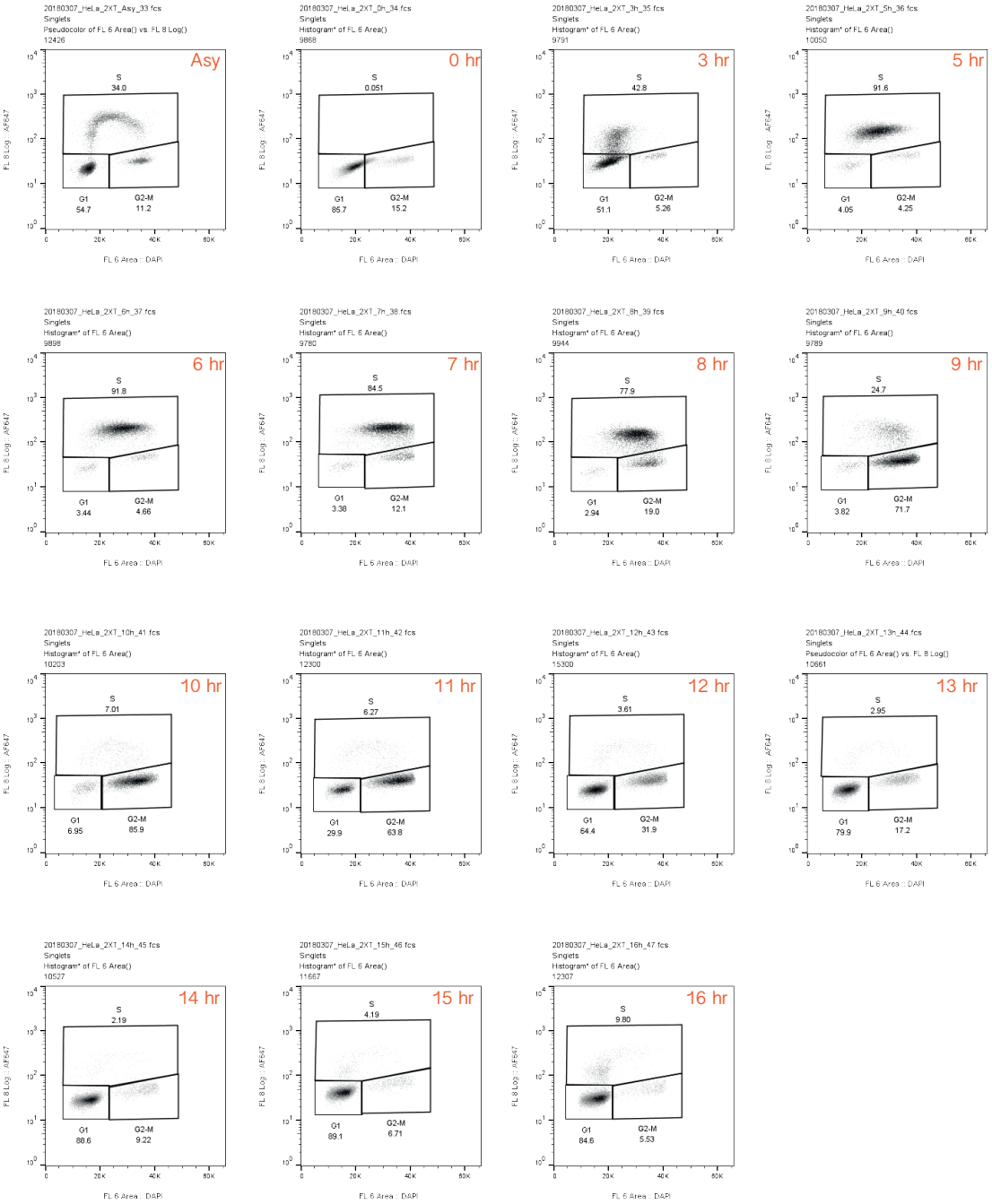

**Supplementary Figure S1 Gating strategy of single cell sorting using the BD AriaIII FACS.** The P4 gate shows the cells selected for plating into the 96-well plate.

```
macro "Projection_Max_blue_red_gray" {
dir1 = getDirectory("Choose Source Directory") ;
dir2 = getDirectory("Choose Destination Directory") ;
list = getFileList(dir1);
        L = 0 ;
        i = 0 ;
        savdir = 0;
        setBatchMode(true);
        for (i=0; i<list.length; i++) {
                showProgress(i+1, list.length);

                Temp = File.getName(dir1+list[i]);
                if ((endsWith(Temp, ".oif")) ||  (endsWith(Temp, ".oib"))) {
                open(dir1+list[i]); N = getTitle();
                }
run("Z Project...", "projection=[Max Intensity]");
run("Blue");
Stack.setChannel(2);
run("Red");
Stack.setChannel(3);
run("Grays");
run("Duplicate...", "duplicate");
run("Make Composite");
run("RGB Color");
saveAs("Tiff", dir2+N+"-1.oib (RGB)");
selectWindow("MAX_"+N);
run("Split Channels");
selectWindow("C1-MAX_"+N);
run("RGB Color");
saveAs("Tiff", dir2+"C1-MAX_"+N);
selectWindow("C2-MAX_"+N);
run("RGB Color");
saveAs("Tiff", dir2+"C2-MAX_"+N);
selectWindow("C3-MAX_"+N);
run("RGB Color");
saveAs("Tiff", dir2+"C3-MAX_"+N);
run("Collect Garbage");
run("Close All");
}
```

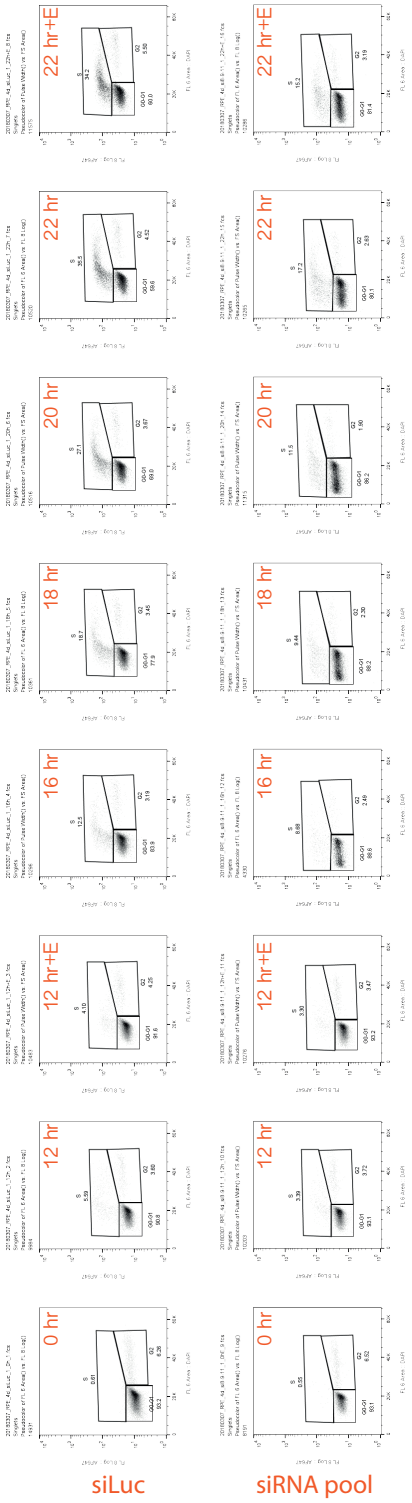
**Supplementary Figure S2 Image projection macro for FIJI.** This macro is written in IJ1 Macro language and assignes the colors blue, red and gray to the respective 3 channels, and creates a .tiff file of the maximum intensity from the z-stacks.

```
macro"Ester[g]"{
	if(roiManager("Count") > 1){
		roiManager("Deselect");
		roiManager("Delete");
	}

	oriID = getImageID();
	run("Z Project...", "projection=[Max Intensity]");
	run("8-bit");

	zproID = getImageID();
	run("Gaussian Blur...", "sigma=2 stack");
	setSlice(1);
	run("Make Binary", "method=Huang background=Dark only");
	run("Invert", "slice");
	run("Fill Holes", "slice");
	run("Watershed", "slice");

	setSlice(3);
	run("Make Binary", "method=Huang background=Dark only");
	run("Invert", "slice");
	run("Watershed", "slice");
	run("Fill Holes", "slice");
	run("Dilate", "slice");
	run("Dilate", "slice");
	run("Dilate", "slice");
	run("Dilate", "slice");

	setSlice(1);
	run("Analyze Particles...", "size=1000-Infinity pixel exclude include add slice");

	roiManager("Remove Channel Info");
	roiManager("Remove Slice Info");

	count = roiManager("count");
	sphase = 0;
	for(i = 0; i < count; i++){
		setSlice(2);
		roiManager("select", i);
		getStatistics(area, mean);
		setResult("Mean",  i, mean);

		setSlice(3);
		roiManager("select", i);
		getStatistics(area, mean);
		if(mean > 240)
			setResult("S-phase",  i, "N");
		else{
			setResult("S-phase",  i, "Y");
			sphase++;
		}
	}
	print("Number of cells in S-phase = ", sphase);
	updateResults();

}
```

**Supplementary Figure S3 Image quantification macro for FIJI.** This macro, written in IJ1 Macro language, determines the area of the nucleus (having DAPI in channel 1) and calculates the intensity of the nuclear signal in channel 2 (in this study $\gamma$H2A.X) in the cells positive or negative for the signal in channel 3 (in the present study EdU).

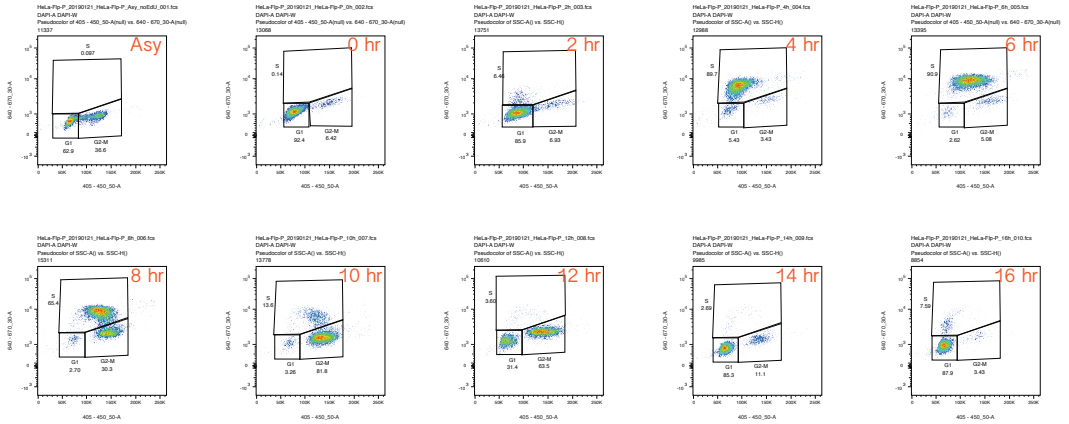



**Supplementary Figure S4 Flow cytometry graphs of EdU-labelled HeLa cells synchronised by DTB.** Orange numbers indicate the time in hours following release from DTB. The x-axis represents the fluorescent intensity of DAPI and the y-axis of Alexa 647-EdU. The percentages of cells in G1, S and G2/M were analysed using the FlowJo software.

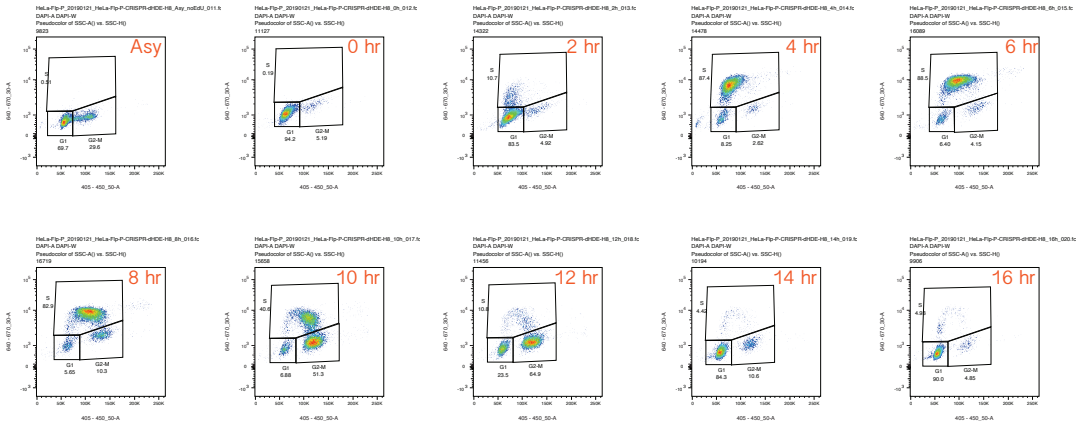



**Supplementary Figure S5 Flow cytometry of EdU-labelled RPE-1 cells depleted of H2A.X polyA RNA and synchronised by contact inhibition.** The left and right columns represent siLuc and siRNA pool transfected cells, respectively. Orange numbers indicate the time in hours after releasing the cells from the $G_0$ arrest. +E: 5 µM etoposide were added 2 hr before harvesting. The x-axis represents the fluorescent intensity of DAPI and the y-axis of Alexa 647-EdU. The percentages of cells in G1, S and G2/M were analysed using the FlowJo software.

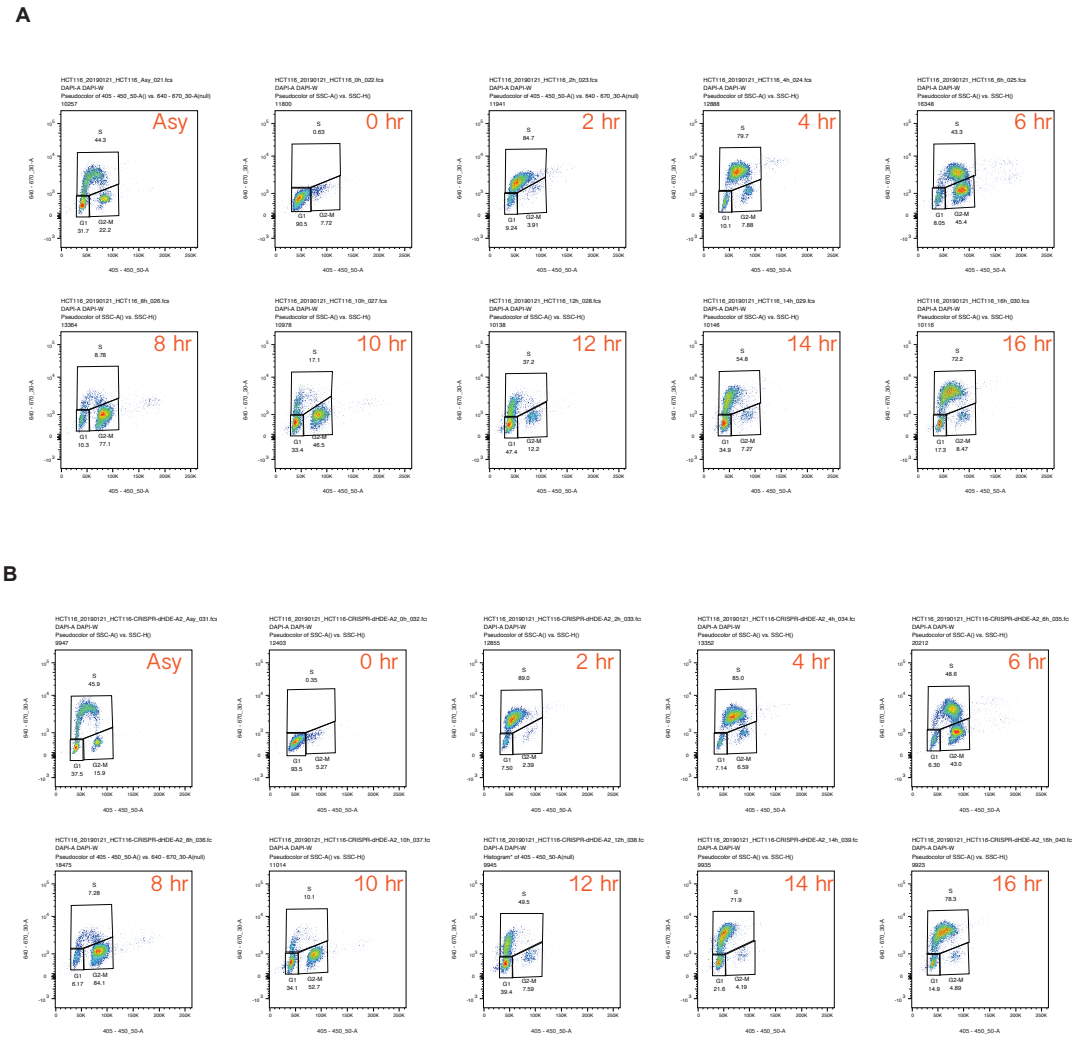
A



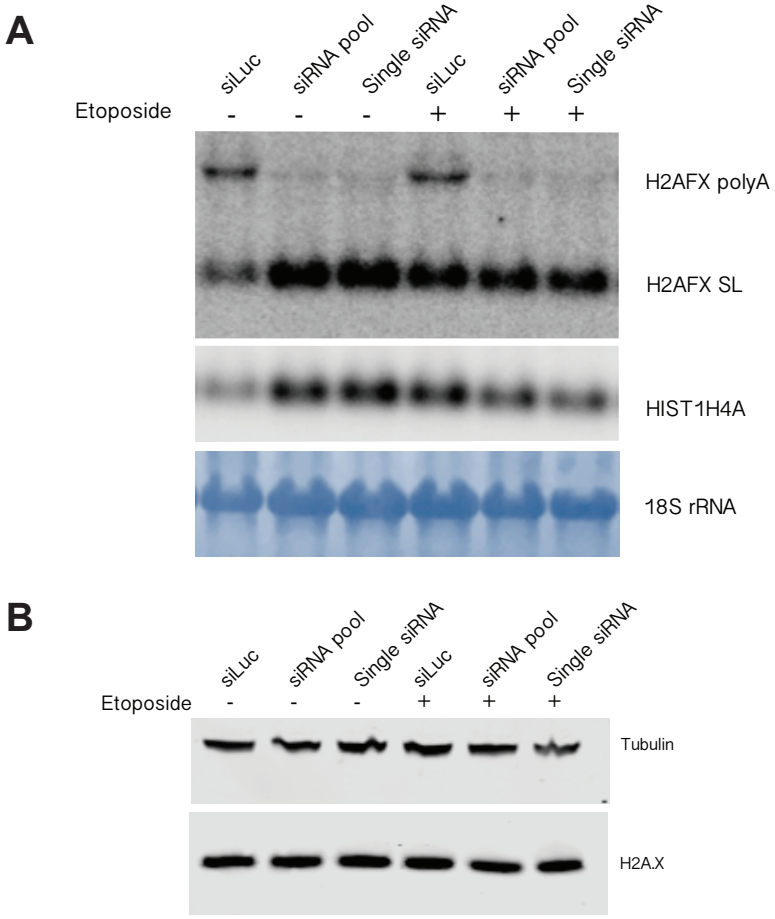

B



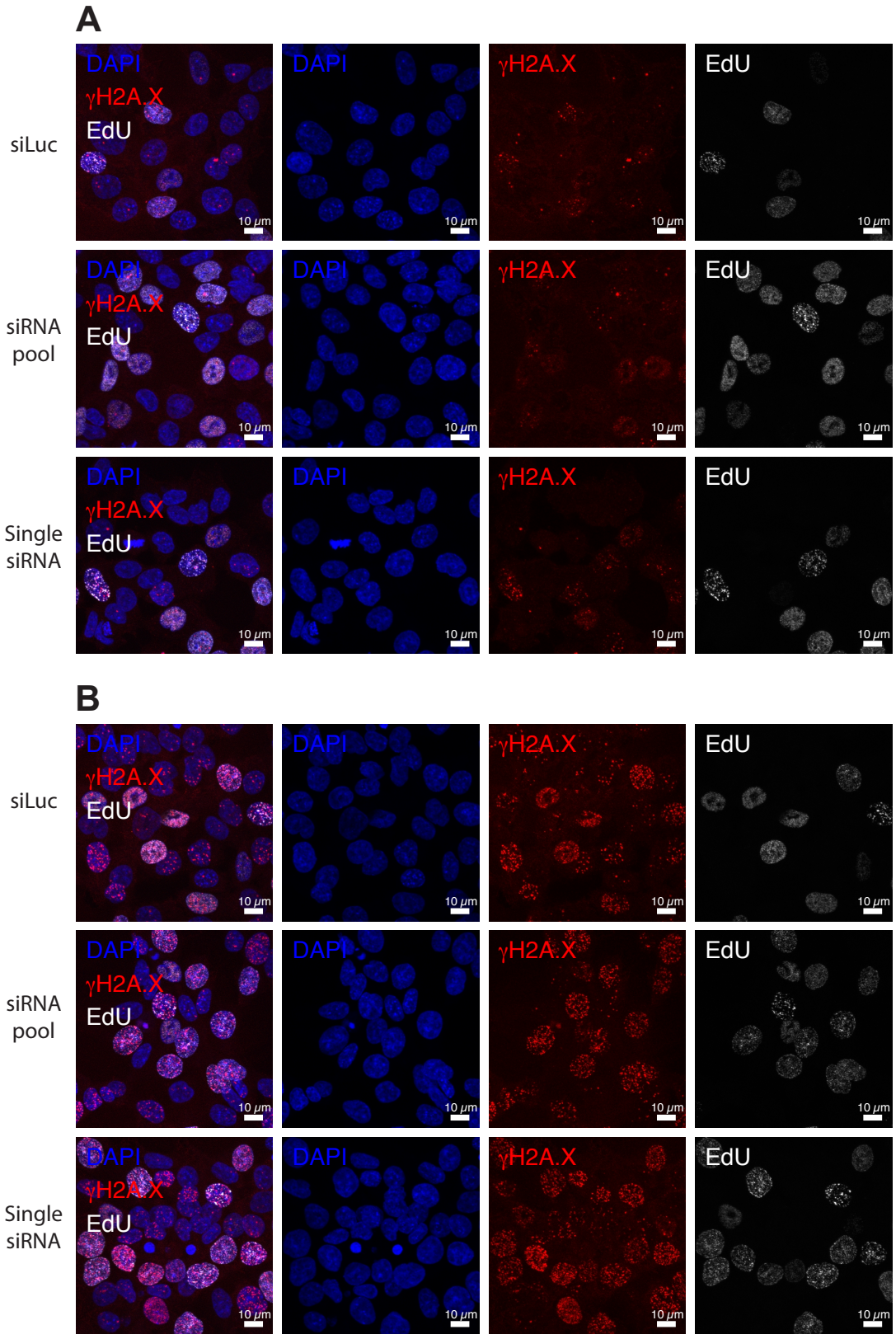

**Supplementary Figure S6 Flow cytometry of EdU-labelled HeLa dHDE CRISPR cells synchronised by DTB.** Orange numbers indicate the time in hours following release from DTB. The x-axis represents the fluorescent intensity of DAPI and the y-axis of Alexa 647-EdU. The percentages of cells in G1, S and G2/M were analysed using the FlowJo software. A) Wild-type HeLa cells. B) HeLa dHDE CRISPR cells.



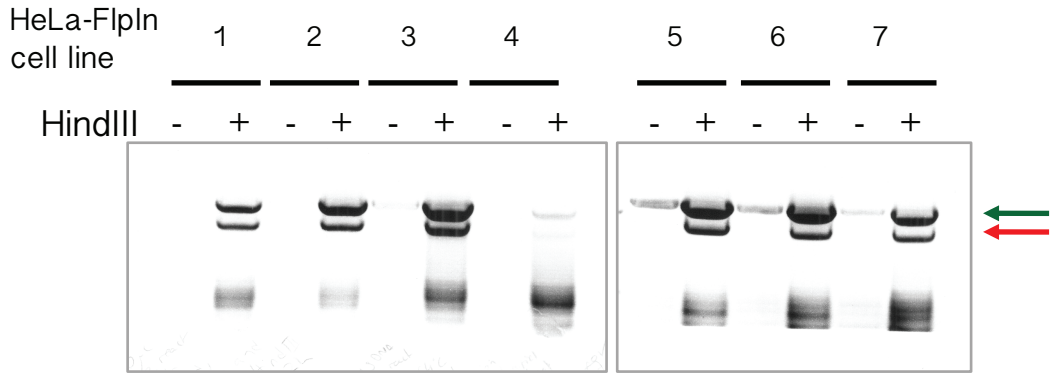

**Supplementary Figure S7 Flow cytometry of EdU-labelled HCT116 dHDE CRISPR cells synchronised by DTB.** Orange numbers indicate the time in hours following release from DTB. The x-axis represents the fluorescent intensity of DAPI and the y-axis of Alexa 647-EdU. The percentages of cells in G1, S and G2/M were analysed using the FlowJo software. A) Wild-type HCT116 cells. B) HCT116 dHDE CRISPR cells.



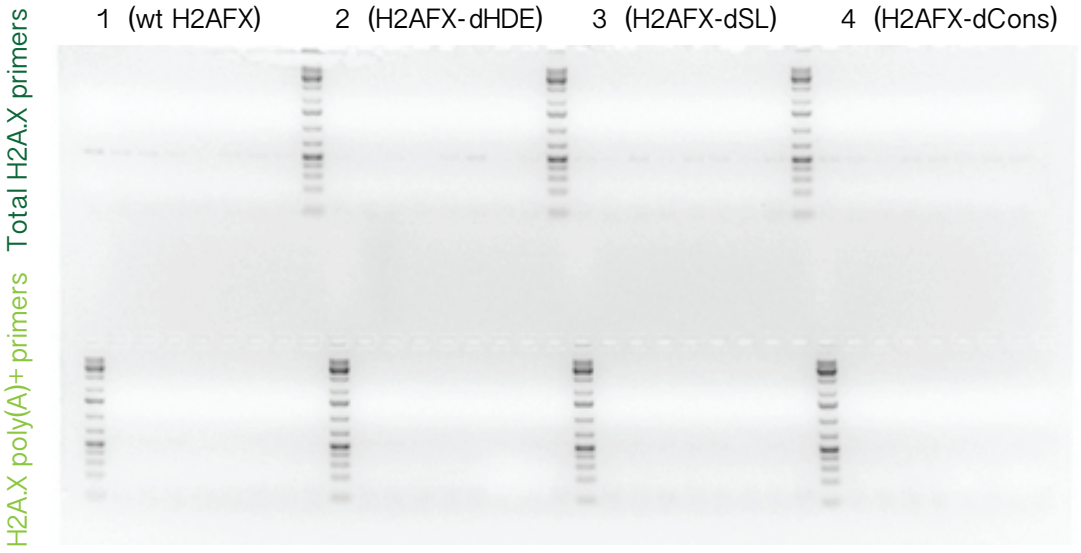

**Supplementary Figure S8 H2A.X poly(A)$^+$ mRNA depletion does not affect total H2A.X protein levels in HCT116 cells.** Wild-type HCT116 cells were transfected with siLuc, siRNA pool or single siRNA for 72 hr. DNA damage was induced with 5 µM etoposide for 2 hr before harvesting. A) Total RNA northern blot. H2A.X SL mRNA and H2A.X poly(A)$^+$ mRNA isoforms were detected with a probe complementary to the common region between both mRNA isoforms. H4 SL: canonical H4 SL mRNA. 18S rRNA: 18S ribosomal RNA (methylene blue staining). Note that in the first lane there is less RNA due to a loading problem. B) Total protein extract western blot for total H2A.X protein levels. Tubulin was used as loading control.



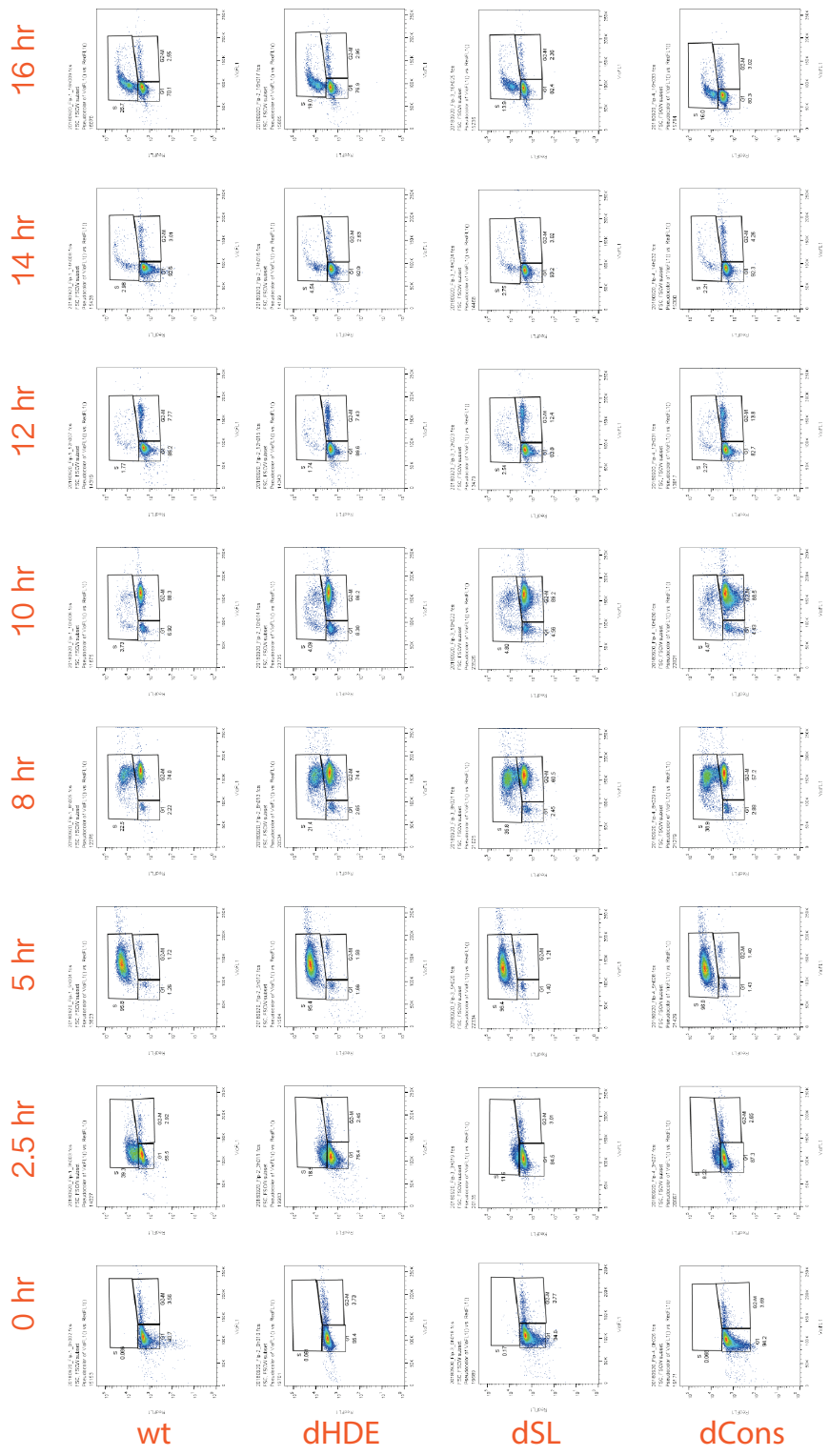

**Supplementary Figure S9 H2A.X poly(A)$^+$ mRNA depletion does not affect γH2A.X signalling in HCT116 cells.** The cells from Supplementary Figure S8 were labelled with EdU (white) to distinguish the S-phase cells, stained with DAPI (blue) to visualise the nuclei and with γH2A.X antibody (red) as DNA damage indicator. A) No etoposide control. B) 5 µM etoposide for 2 hr before harvesting.



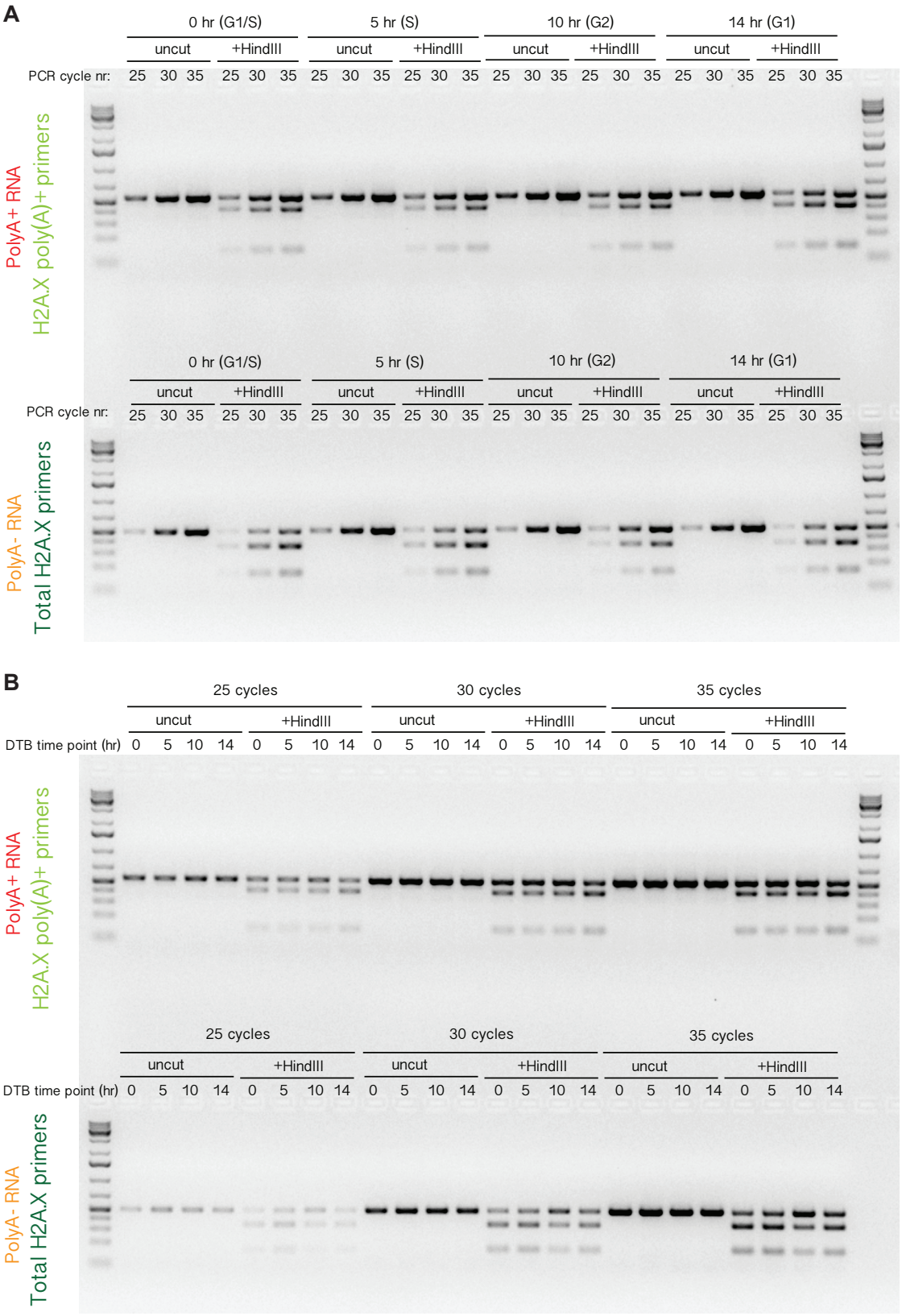

**Supplementary Figure S10 HeLa cells have two H2AFX genes.** Genomic DNA (gDNA) was extracted from the different HeLa-FlpIn cell lines (the numbers correspond to the cell line names shown in Figure 6.1), H2AFX was amplified by PCR and the PCR product was digested with HindIII. Green arrow: band corresponding to the endogenous H2AFX genes (HindIII-resistant). Red arrow: transgene H2AFX (HindIII-sensitive). Note the 2:1 endogenous:transgene ratio. The PCR for cell line #4 was inefficient but the ratio 2:1 endogenous:transgene is consistent.




**Supplementary Figure S11 Minus RT PCR of the HeLa-FlpIn cell lines synchronised by DTB.** The numbers correspond to the constructs shown in Figure 6.1). Reverse transcription reactions were prepared with total RNA from the HeLa-FlpIn cell lines synchronised by DTB, without reverse transcriptase and a PCR was performed to test the presence of gDNA.




**Supplementary Figure S12 Flow cytometry of EdU-labelled HeLa-FlpIn cell lines synchronised by DTB** Orange numbers on the left indicate the time in hours following release from DTB. wt: HeLa-FlpIn cell line stably expressing H2A.X-HindIII. dHDE: HeLa-FlpIn cell line stably expressing H2A.X-HindIII-dHDE. dSL: HeLa-FlpIn cell line stably expressing H2A.X-HindIII-dSL. dCons: HeLa-FlpIn cell line stably expressing H2A.X-HindIII-dCons. The x-axis represents the fluorescent intensity of DAPI and the y-axis of Alexa 647-EdU. The percentages of cells in G1, S and G2/M were analysed using the FlowJo software.




**Supplementary Figure S13 The PCR cycle number does not affect the transgene/endogenous ratio.**