Biochemical and functional characterization of the RNA binding protein SAF-A/hnRNP U and investigation of its potential role in mammalian X chromosome inactivation

Anna Tattermusch

Somerville College

Dissertation submitted for the degree of Doctor of Philosophy

Department of Biochemistry

University of Oxford

August 2015

Abstract

Biochemical and functional characterization of the RNA binding protein SAF-A/hnRNP U and investigation of its potential role in mammalian X chromosome inactivation

Anna Tattermusch

Somerville College

Submitted for the degree of Doctor of Philosophy, October 2014

Resubmitted with corrections, August 2015

X chromosome inactivation (XCI) is an important mechanism for gene dosage compensation between XY males and XX females during early embryonic development of eutherian mammals. The major regulator of XCI is the long non-coding RNA (lncRNA) X inactive specific transcript (Xist). Xist RNA is believed to recruit epigenetic modifiers and initiate a silencing cascade that alters the chromatin structure and culminates in the transcriptional repression of most genes at the inactive X chromosome (Xi). Though extensively studied, the protein factors involved in the recruitment, spreading and maintenance of Xist RNA at the Xi are largely unknown.

Recently published studies provide evidence that the DNA and RNA binding nuclear matrix protein scaffold attachment factor A/heterogeneous ribonucleoprotein U (SAF-A/hnRNP U) is required for Xist localization to the Xi. These studies, however, are incomplete in that they neither provide information on the RNA binding consensus sequence for SAF-A/hnRNP U on Xist, nor on its functional role in XCI.

The present study was designed to re-investigate the role of SAF-A/hnRNP U in XCI using state-of-the-art techniques, notably fluorescence recovery after photo-bleaching (FRAP), 3D-structured illumination microscopy (3D-SIM), and *in vivo* crosslinking and RNA immunoprecipitation (CLIP).

The results indicate that SAF-A/hnRNP U exists in two forms; a highly dynamic pan-nuclear protein and a stably Xi-associated form. 3D-SIM reveals that the latter closely associates with Xist RNA, and lines out the interchromatin channels within the Barr body. This finding suggests that SAF-A/hnRNP U takes part in the structural organization of the Xi by creating a secluded chromatin environment that retains XCI factors in their dedicated nuclear location. In contrast to observations made in microscopy imaging, biochemical experiments did not confirm a direct binding of SAF-A/hnRNP U to Xist RNA. This observation led to the conclusion that the protein plays an indirect, structural role in XCI.

List of publications

1. Nesterova TB, Senner CE, Schneider J, Alcayna-Stevens T, **Tattermusch A**, Hemberger M, Brockdorff N, *Pluripotency factor binding and Tsix expression act synergistically to repress Xist in undifferentiated embryonic stem cells.* Epigenetics Chromatin, 2011. 4(1): p. 17.

2. **Tattermusch A**, Brockdorff N, *A scaffold for X chromosome inactivation.* Hum Genet, 2011. 130(2): p. 247-53.

3. Cerase A, Smeets D, Tang AY, Gdula M, Kraus F, Spivakov M, Moindrot B, Leleu M, **Tattermusch A**, Demmerle J, Nesterova TB, Green C, Otte AP, Schermelleh L, Brockdorff N, *Spatial separation of Xist-RNA and polycomb proteins revealed by super-resolution microscopy.* Proc Natl Acad Sci U S A, 2014. 111(6): p. 2235-40.

4. Smeets D, Maraki Y, Schmid VJ, Kraus F, **Tattermusch A**, Cerase A, Sterr M, Fiedler S, Demmerle J, Popken J, Leonhardt H, Brockdorff N, Cremer T, Schermelleh L, *Three-dimensional super-resolution microscopy of the inactive X chromosome territory reveals a collapse of its active nuclear compartment harboring distinct Xist RNA foci.* Epigenetics Chromatin, 2014. 7: p. 8.

5. Cerase A, Pintacuda G, **Tattermusch A**, Avner P, *Xist localization and function: new insights from multiple levels.* Review article accepted by Genome Biology.

Declaration of authorship

The thesis submitted is entirely my own work, except where otherwise indicated, and has been completed between January 2011 and October 2014, and corrections have been made between the end of February and the beginning of August 2015. I have clearly indicated the presence of material I have quoted from other sources. This thesis has not been submitted, either partially or in full, for another qualification of this University or for a qualification at any other institution.

Anna Tattermusch

August 2015

Acknowledgements

I would like to thank Dr. Neil Blackledge (Department of Biochemistry, University of Oxford, UK) for the pNIC28-Bsa4 vector, Dr. Anton Wutz (Wellcome Trust Centre for Stem Cell Research Cambridge, UK) for the pCAGGS_mSAF-A:hrGFP construct, Dr. Sanichi Nakagawa (RIKEN Wako, Japan) for the Neuro2a cell lines, Dr. Andrea Cerase (Department of Biochemistry, University of Oxford, UK) for the 3E and P6D7 cells, immuno-fluorescence *in situ* hybridization (immuno-FISH) preparations and helpful discussions. Special thanks go to Daniel Smeets (LMU Munich, Germany), Dr. Lothar Schermelleh (Department of Biochemistry, University of Oxford, UK), Dr. David Pinto (Department of Biochemistry, University of Oxford, UK), Dr. med. Marion Cremer (LMU Munich, Germany) and Prof. Dr. med. Thomas Cremer (LMU Munich, Germany) for the high-resolution imaging. I would also like to thank Dr. Alexander Kurze (Department of Biochemistry, University of Oxford, UK) and Dr. Martin Houlard (Department of Biochemistry, University of Oxford, UK) for help with the fluorescence recovery after photobleaching (FRAP) imaging and the long-term time-lapse imaging, respectively. Furthermore, I would like to acknowledge Dr. Nojima Takayuki (Sir William Dunn School of Pathology, University of Oxford, UK) for advice and help with the initial *in vivo* UV cross-linking and immunoprecipitation (CLIP) experiment, and Dr. Lorna Gregory, Dr. Paolo Piazza and Dr. Amy Trebes (The Wellcome Trust Centre for Human Genetics, University of Oxford, UK) for advice on experimental design and CLIP cDNA library preparations and high-throughput sequencing (HTS), and I am grateful to Dr. Russell Hamilton (Department of Biochemistry, University of Oxford, UK) for making attempts to help me with the initial analysis of my CLIP data. I would also like to thank Dr. Benjamin Thomas and Svenja Hester (Central Proteomics Facility, Sir William Dunn School of Pathology, University of Oxford) for help and advice with mass spectrometry and Dr. David Trudgian (UT Southwestern Medical Center, Dallas) assistance with data analysis. Moreover, I would like to thank Anne Hansen and Dr. Katja Simon (Weatherall Institute of Molecular Medicine, Human Immunology Unit, John Radcliffe Hospital, Oxford, UK) for their help with the Image

Stream IS100 experimental design, setup and data analysis. Lastly, I would like to thank the German Business Foundation (SDW) and the Medical Research Council (MRC) and Prof. Dr. Neil Brockdorff for funding me and this work.

Additionally, I would like to thank all those who have accompanied me on this way. At this point I would like to express my gratitude to my family for their support and love, Edgar for bringing passion into my life, and my dancing partner Felix for companionship. Furthermore, I would like to thank my dancing coach and mentor Neil for his patience, and all my other friends for their constant presence and attention.

Errare humanum est.

Abbreviations

°C	degree Celsius
3D-SIM	3D-structured illumination microscopy
4SU	4-thiouridine
aa	amino acids
aRRM	atypical RNA recognition motif
ASF	Alternative Splicing Factor
Ash2l	Absent, small or homeotic discs 2-like
ATP	adenosine triphosphate
AUKB	Aurora Kinase B
BLAST	Basic Local Alignment Search Tool
bp	base pair
BSA	bovine serum albumin
BUR	base unpairing region
CBB	Coomassie Brilliant Blue G-250
CHART-seq	capture hybridization analysis of RNA targets with deep sequencing
cDNA	complementary DNA
cHC	constitutive HC
ChIP	chromatin IP
CLIP	*in vivo* crosslinking and IP
cM	centimorgan
CRD	coding region instability determinant
CRISPR	clustered regularly interspaced short palindromic repeat
CT	carboxy-terminal
Da	Dalton

DAPI	4',6-diamidino-2-phenylindole
DIC	differential interference contrast (microscopy)
DNA	deoxyribonucleic acid
DNA-PK	DNA-protein kinase complex
dox	doxycycline
dpc	*days post coitum*
EC	euchromatin
EdU	5-ethynyl-2'-deoxyuridine
EED	Embryonic ectoderm development
eGFP	enhanced GFP
ESC	embryonic stem cell
Ezh1/2	Enhancer of zeste 1/2
FA	formaldehyde
FACS	fluorescence-activated cell sorting
FCS	fluorescence correlation spectroscopy
fHC	facultative HC
FISH	fluorescence *in situ* hybridization
FL	full-length
FPALM	fluorescence photoactivation localization microscopy
FRAP	fluorescence recovery after photobleaching
g	gravitational force
G1 phase	gap 1 phase
GFP	Green Fluorescent Protein
GO	gene ontology
H	Histone
h	hour(s)

HAVANA	Human and Vertebrate Analysis and Annotation
HC	heterochromatin
HEK	human embryonic kidney
hnRNA	heterogeneous nuclear RNA
hnRNP	heterogeneous ribonucleoproteins
hrGFP	humanized *Renilla reniformis* GFP
HTS	high-throughput sequencing
ICM	inner cell mass
ID	identifier
iCLIP	individual-nucleotide resolution CLIP
IF	immunofluorescence
iFRAP	inverse FRAP
IP	immunoprecipitation
K	lysine
KD	knockdown
kDa	kilo Dalton
KH	hnRNP K homology domain
KO	knockout
LCSM	laser scanning microscopy
LIC	ligation-independent cloning
LINE	long interspersed element
lncRNA	long non-coding RNA
LTR	long terminal repeat
me1	monomethyl
me2	dimethyl
me3	trimethyl
MeAc	methanol acetone

MeOH	methanol
min	minute(s)
mm	*Mus musculus*
mRNA	messenger RNA
MS	mass spectrometry
MSCI	meiotic sex chromosome inactivation
MW	molecular weight
n	nano
NCBI	National Center for Biotechnology Information
nd	data not available
ne	nuclear extract
NT	amino-terminal
OTF	optical transfer function
PAGE	polyacrylamide gel electrophoresis
PALM	photoactivated localization microscopy
PAR-CLIP	Photoactivatable Ribonucleoside-enhanced CLIP
PcG	Polycomb Group
PCGF	Polycomb Group Ring Finger
PFA	paraformaldehyde
pFSPE	parental FSPE
PI	propidium iodide
PP1	Protein Phosphatase 1
PRC1/2	Polycomb Repressive Complex 1/2
PRMT1	protein arginine methyltransferase 1
Psc	Posterior sex combs
RAP	RNA antisense purification
RBD	RNA binding domain

RBBP4/7	Retinoblastoma binding protein 4/7
RepA	A-repeat (of Xist RNA)
RING1A/B	Really interesting gene 1A/1B
RING1B	Really Interesting New Gene 1B
RNA	ribonucleic acid
RNAi	RNA interference
RRM	RNA recognition motif
rRNA	ribosomal RNA
s/sec	second(s)
SET	Su(var)3-9, Enhancer-of-zeste, Trithorax
SINE	short interspersed element
S phase	synthesis phase
S/MAR	scaffold/matrix attachment region
SAF-A/hnRNP U	scaffold attachment factor A/heterogeneous ribonucleoprotein U
SATB1	Special AT-rich Binding 1
SDS	sodium dodecyl sulfate
SELEX	Systematic Evolution of Ligands by Exponential Enrichment
SF2	Splicing Factor-2
SINQ	spectral index quantification
siRNA	small interfering RNA
SmcHD1	Structural maintenance of chromosomes flexible Hinge Domain containing 1
SNP	single nucleotide polymorphism
spf	*sparse fur*
SPRY domain	spIA/Ryanodine receptor domain

ssDNA	single-stranded DNA
STED	stimulated emission depletion microscopy
STORM	stochastic optical reconstruction microscopy
Suz12	Suppressor of zeste 12
TrxG	Trithorax Group
TSC	trophoblast stem cell
U	ubiquitin
UV	ultraviolet
WB	western blot
wce	whole cell extract
wt	wild type
Xa	active X chromosome
Xce	X controlling element
XCI	X chromosome inactivation
Xi	inactive X chromosome
Xic	X inactivation center
Xist	X inactive specific transcript
Xm	maternal X
Xp	paternal X
YY1	Ying Yang 1
µ	micro

Table of contents

Abstract 2

List of publications 4

Declaration of authorship 5

Acknowledgements 6

Abbreviations 8

Table of contents 14

List of figures 20

List of tables 24

1. Introduction 25

1.1 An introduction to X chromosome inactivation 25

1.1.1 A brief history of X inactivation 25

1.1.2 X chromosome inactivation in eutherian mammals 25

1.1.3 Escape from X inactivation 26

1.1.4 X chromosome inactivation during mouse embryonic development 28

1.2 The inactive X chromosome and Xist RNA 31

1.2.1 Higher-order chromatin structure at the female X chromosomes 31

1.2.2 The X inactivation center and the discovery of Xist RNA 34

1.2.3 Polycomb group proteins in X chromosome inactivation 37

1.2.4 Protein factors involved in X chromosome inactivation 38

1.2.5 Xist RNA spreading and LINE repeats 41

1.2.6 Xist RNA localization during the cell cycle 44

1.3 The nuclear matrix/scaffold – foundation for nuclear organization 46

1.3.1 The nuclear matrix/scaffold is a controversial structure 48

1.3.2 Heterogeneous ribonucleoproteins – Eclectic RNA binding proteins 49

1.3.3 SAF-A/hnRNP U – A DNA and RNA binding nuclear matrix protein 51

1.3.4 SAF-A/hnRNP U in X chromosome inactivation 53

1.4 Aims of this thesis 55

2. SAF-A/hnRNP U is a pan-nuclear scaffolding protein 57

2.1 Introduction 57

2.2 Generation of an antibody specifically recognizing SAF-A/hnRNP U 59

2.3 SAF-A/hnRNP U associates with other RNA binding factors 65

2.4 SAF-A/hnRNP U is an extensively ubiquitinated protein 75

2.5 SAF-A/hnRNP U phosphorylation is undetectable 76

2.6 SAF-A/hnRNP U is a nuclear protein with broad nucleoplasmic staining pattern 79

2.7 Summary and discussion 83

2.7.1 SAF-A/hnRNP U protein interactions 83

2.7.2 SAF-A/hnRNP U post-translational modifications 84

2.7.3 SAF-A/hnRNP U nuclear localization and visualization at the Xi 85

3. *In vivo* UV crosslinking and immunoprecipitation to characterize SAF-A/hnRNP U RNA binding activity 88

3.1 Introduction 88

3.2 SAF-A/hnRNP U CLIP - Preparatory experiments 91

3.3 SAF-A/hnRNP U CLIP experiment 96

3.4 SAF-A/hnRNP U CLIP-Seq data quality control 99

3.5 Qualitative SAF-A/hnRNP U CLIP-Seq data analysis 117

3.6 Summary and discussion..... 120

3.6.1 CLIP-Seq experimental considerations..... 120

3.6.2 SAF-A/hnRNP U RNA binding behavior..... 123

4. SAF-A/hnRNP U stably associates with the Xi 126

4.1 Introduction..... 126

4.2 hrGFP-tagged SAF-A/hnRNP U is enriched at the Xi..... 126

4.3 SAF-A/hnRNP U:hrGFP stably associates with the Xi 129

4.4 Summary and discussion..... 133

4.4.1 SAF-A/hnRNP U:hrGFP localization in FSPE cells 133

4.4.2 SAF-A/hnRNP U:hrGFP dynamics at the Xi and in the nuclear surroundings 135

5. SAF-A/hnRNP U does not associate with the Xi during mitosis..... 137

5.1 Introduction..... 137

5.2 FSPE cell cycle analysis..... 138

5.3 SAF-A/hnRNP U and Xist RNA cell cycle dynamics are different..... 141

5.4 The effect of phosphatase and kinase inhibitors on SAF-A/hnRNP U localization.... 143

5.5 SAF-A/hnRNP U *in vivo* cell cycle dynamics..... 147

5.6 PCGF3 and SAF-A/hnRNP U cell cycle dynamics are similar..... 150

5.7 Preliminary FACS cell cycle and imaging analysis on SAF-A/hnRNP U and PCGF3154

5.8 Summary and discussion..... 160

5.8.1 SAF-A/hnRNP U release precedes Xist RNA dissociation from the Xi 160

5.8.2 Influence of tautomycin and hesperadin on SAF-A/hnRNP U localization..... 161

5.8.3 SAF-A/hnRNP U and PCGF3 cell cycle dynamics are comparable 162

6. Endogenous and hrGFP-tagged SAF-A/hnRNP U in super-resolution imaging 163

6.1 Introduction 163

6.2 SAF-A/hnRNP U associates with Xist RNA 167

6.3 Summary and discussion 173

6.3.1 SAF-A/hnRNP U associate with Xist RNA 173

7. General discussion - A role for SAF-A/hnRNP U in mouse XCI? 175

7.1 SAF-A/hnRNP U nuclear localization 175

7.2 SAF-A/hnRNP U dynamics 176

7.3 SAF-A/hnRNP U RNA binding 177

7.4 Additional considerations 181

7.4.1 Genetic manipulation of SAF-A/hnRNP U or Xist RNA expression 181

7.4.2 SAF-A/hnRNP U chromatin immunoprecipitation 182

8. Materials and Methods 184

Abbreviations 184

8.1 Molecular cloning 186

8.1.1 PCR reaction setup 187

8.1.2 LIC cloning of N-terminal murine SAF-A/hnRNP U 187

8.1.3 TOPO cloning 187

8.2 Protein expression and purification 188

8.2.1 Purification of six Histidine-tagged proteins 188

8.3 Antibody preparation and characterization 189

8.3.1 Antibody production 189

8.3.2 Antibody affinity purification 189

8.3.3 Antibody-antigen competition assay 190

8.4 Antibodies 190

8.5 Protein biochemistry 191

8.5.1 Whole cell extract preparation 191

8.5.2 Nuclear extract preparation 191

8.5.3 *In situ* nuclear matrix preparation 192

8.5.4 SDS-PAGE 192

8.5.5 Western blotting 193

8.5.6 Immunoprecipitation 193

8.5.7 *In vivo* UV crosslinking and immunoprecipitation 194

8.5.8 CLIP data analysis 195

8.5.9 Mass spectrometry 195

8.5.10 Phos-tag™ sample and gel preparation 196

8.6 General molecular biology techniques 197

8.6.1 RNA extraction 197

8.6.2 Complementary DNA synthesis 197

8.6.3 Quantitative real-time PCR 198

8.7 Tissue culture 198

8.7.1 Cell culture media 198

8.7.2 Freezing and thawing cells 199

8.7.3 Feeder cell preparation 199

8.7.4 Lipofection of plasmid DNA 200

8.7.5 Kill curve preparation 200

8.7.6 Cell lines 201

8.7.7 Stable cell line preparation and colony picking 202

8.8 Microscopy techniques 203

8.8.1 Indirect immunofluorescence 203

8.8.2 Immuno-FISH for 3D-SIM and conventional fluorescence microscopy 204

8.8.3 3D-SIM microscopy 205

8.8.4 FRAP and iFRAP 206

8.8.5 Long-term time-lapse imaging 206

8.8.6 Fiji image analysis 207

8.9 Flow cytometry 207

8.9.1 Propidium iodide staining for FACS 207

8.9.2 DAPI staining for ImageStream IS100 207

8.10 Figure preparation 208

9. References 209

List of figures

Figure 1.1 – XCI during mouse embryonic development. 30

Figure 1.2 - Nucleosome structure and chromatin. 32

Figure 1.3 – Human and mouse XIST/Xist RNA gene structures..... 36

Figure 1.4 – Gartler and Riggs model for the propagation of XCI. 42

Figure 1.5 – Xist RNA interphase localization. 45

Figure 1.6 – The nuclear matrix in electron microscopy. 47

Figure 1.7 – Schematic domain structures of major hnNRPs. 50

Figure 1.8 – Model for SAF-A/hnRNP U action at the Xi proposed by Hasegawa *et al.* (2010). 54

Figure 2.1 – Antigen design and purification. 59

Figure 2.2 – Anti-SAF-A/hnRNP U recognizes the recombinant fragment and endogenous protein. 60

Figure 2.3 – Ensembl database entries for mouse and human SAF-A/hnRNP U. 62

Figure 2.4 – Affinity purification further reduces anti-SAF-A/hnRNP U background signal... 63

Figure 2.5 – SAF-A/hnRNP U immunoprecipitation. 64

Figure 2.6 – Confirmation of SAF-A/hnRNP U specificity by mass spectrometry. 65

Figure 2.7 – Gene ontology analysis for the SAF-A/hnRNP U interactome. 67

Figure 2.8 – Quality control showing suitability of power law global error model (plgem) analysis for the normalized spectral index quantification (SINQ) data. 70

Figure 2.9 – SAF-/hnRNP U interactome analysis shows preferential association with nucleotide binding proteins..... 71

Figure 2.10 – Quality control showing low quality of SINQ IP data in the presence of RNase. 74

Figure 2.111 – SAF-A/hnRNP U appears unphosphorylated in Phos-tag SDS-PAGE..... 78

Figure 2.122 – Anti-SAF-A/hnRNP U specifically recognizes the SAF-A/hnRNP U epitope in IF. 80

Figure 2.133 – Anti-SAF-A/hnRNP U does not show the enrichment at the Xi using different fixation techniques. 81

Figure 2.14 – Anti-SAF-A/hnRNP U does not show the enrichment on the Xi in IF on a variety of cell lines..... 82

Figure 2.15 – Endogenous SAF-A/hnRNP U enrichment at the Xi cannot be visualized in *in situ* nuclear matrix preparation. 83

Figure 2.166 – Models for anti-SAF-A/hnRNP U epitope exclusion at the Xi. 87

Figure 3.1 – Primary amino acid sequence of murine SAF-A/hnRNP U. 89

Figure 3.2 – Workflow for an *in vivo* UV crosslinking and immunoprecipitation high-throughput sequencing experiment (CLIP-Seq)..... 91

Figure 3.3 – Characterization of a male 3E ESC line carrying a doxycycline-inducible autosomal Xist transgene..... 93

Figure 3.4 – SAF-A/hnRNP U protein in nuclear extract preparation fractions..... 94

Figure 3.5 – Xist content of nuclear extract preparations..... 95

Figure 3.6 – SAF-A/hnRNP U CLIP autoradiography. 97

Figure 3.7 – SAF-A/hnRNP U CLIP cDNA library preparation and initial quality control. 98

Figure 3.8 – Per base pair sequence quality of SAF-A/hnRNP U CLIP sample and E14 input controls (libraries 1 and 2)..... 102

Figure 3.9 – Per sequence quality of SAF-A/hnRNP U CLIP sample and E14 input controls (libraries 1 and 2). 103

Figure 3.10 – Per base sequence content of SAF-A/hnRNP U CLIP sample and E14 input controls (libraries 1 and 2). 106

Figure 3.11 – Per base GC content in the SAF-A/hnRNP U CLIP library and E14 input controls (libraries 1 and 2). 107

Figure 3.12 – Per sequence GC content of SAF-A/hnRNP U CLIP library and E14 input controls (libraries 1 and 2). 110

Figure 3.13 – Sequence duplication levels in the SAF-A/hnRNP U CLIP library and E14 input controls (libraries 1 and 2). 111

Figure 3.14 – Kmer profiles for SAF-A/hnRNP U CLIP library and E14 input controls (libraries 1 and 2). 115

Figure 3.15 – IGV profiles of mouse ESC input controls and the SAF-A/hnRNP U sample. 118

Figure 4.1 – SAF-A/hnRNP U:hrGFP stable cell line preparation. 127

Figure 4.2 – Co-localization of Xist RNA and Xi-enriched hrGFP-tagged SAF-A/hnRNP U. 129

Figure 4.3 – Stable association of SAF-A/hnRNP U with the Xi. 131

Figure 4.4 – Exemplary SAF-A/hnRNP U FRAP image series. 132

Figure 4.5 – Endogenous and FLAG-tagged SAF-A/hnRNP U in Neuro2a cells. 134

Figure 4.6 – SAF-A/hnRNP U FRAP summary. 136

Figure 5.1 – Schematics of a typical eukaryotic cell cycle and FSPE cell cycle profile. 140

Figure 5.2 – Relative localization of SAF-A/hnRNP U and Xist RNA in FSPEs throughout the cell cycle. 142

Figure 5.3 – Tautomycin and hesperadin: two drugs altering XIST-Xi associations. 144

Figure 5. 4 – Drug treatment and SAF-A/hnRNP U localization in FSPE cells. 146

Figure 5.5 – Western blotting screen of SAF-A/hnRNP U:hrGFP – H2B:mCherry stable FSPE cell lines. 147

Figure 5.6 – SAF-A/hnRNP U dynamics in single cell time-lapse imaging. 150

Figure 5.7 – Screen for Xi-association of PRC1 proteins in FSPE cells in fixed – and live – cell imaging. 152

Figure 5.8 – PCGF3 dynamics in single cell, time-lapse imaging. 153

Figure 5.9 – FACS images of interphase SAF-A/hnRNP U:hrGFP. 155

Figure 5.10 – Preliminary FACS gating strategies for cell cycle stages and PCGF3 at the Xi. 157

Figure 5.11 – FACS images of cell cycle sorted PCGF3:GFP events. 159

Figure 6.1 – Co-localization in and resolution of 3D-SIM. 164

Figure 6.2 – Enhanced image resolution through the moiré effect. 165

Figure 6.3 – 3D-SIM alignment and secondary antibody controls. 168

Figure 6.4 – 3D-SIM analysis workflow and example images. 171

Figure 6.5 – Spatial overlap of Xist RNA and endogenous SAF-A/hnRNP U at C2C12 Barr bodies. 172

Figure 7.1 – Models for SAF-A/hnRNP U molecular associations. 179

Figure 7.2 – SAF-A/hnRNP U confines Xist RNA to the Xi territory. 180

Figure 8.1 – Effect of Geneticin G418 on FSPE cell viability. 201

List of tables

Table 2.1 – Proteins enriched and depleted in SAF-A/hnRNP U co-IPs. 72

Table 2.2 – SAF-A/hnRNP U ubiquitination sites identified by error tolerant Mascot searches. 76

Table 3.1 – CLIP-Seq RNA read alignment overview. 100

Table 3.2 – List of overrepresented sequences in the SAF-A/hnRNP U CLIP library. 113

Table 8.1 – Antibodies. 190

Table 8.2 – qPCR primes sets. 198

Table 8.3 – Cell lines..... 201

1. Introduction

1.1 An introduction to X chromosome inactivation

1.1.1 A brief history of X inactivation

In 1949 Murray Barr published a report on the observable differences between nuclei in male and female feline neurons (Barr and Bertram 1949). In his article Barr described that "something distinct" in the chromatin content of these cells determined their femaleness. Under the microscope he identified small, dark sub-nuclear structures adjacent to the nucleolus that were unique to female nuclei. Later on, the Japanese cell biologist Susumu Ohno identified this enigmatic, female-specific structure as part of the sex chromatin (Ohno, Kaplan et al. 1959). Mary Lyon then managed to put together all the pieces of the jigsaw, and proposed the Lyon hypothesis, which nowadays is also known as Lyon's law, stating that one of the two Xs in female cells is inactivated (Lyon 1961).

What distinguishes this particular X chromosome from its homologous partner? - Upon closer examination it was found that the X, discovered by Barr, lacks transcriptional activity. Consequently, it was termed inactive X chromosome (Xi). Notably, the Xi is often referred to as the Barr body. Conversely its transcriptionally active, homologous partner was called the active X chromosome (Xa). The Lyon hypothesis summarizes all these findings, stating that female mammals are mosaics in terms of sex chromatin expression. During the years after the proposal of Lyon's hypothesis major efforts have been made to elucidate the molecular mechanisms behind this biological phenomenon.

1.1.2 X chromosome inactivation in eutherian mammals

Owing to divergent evolutionary processes the eutherian X and Y chromosomes are dissimilar in terms of size and gene content, with the X chromosome being larger and enriched in genes as compared to the small and gene-depleted Y chromosome (Ross, Grafham et al. 2005).

In course of the past 300 million years, the two types of mammalian sex chromosomes have evolved from a pair of autosomes, one of which carried a sex-determining gene (Muller 1964, Charlesworth 1978, Jablonka and Lamb 1990, Charlesworth 1991). Over time, one of these chromosomes lost almost all traces of the ancestral autosome and evolved into what is now the Y chromosome. The X, on the contrary, maintained many of the autosomal features (Muller 1964, Charlesworth 1978, Rice 1987, Charlesworth 1991). Consequently, the loss of genetic activity from the evolving Y chromosome resulted in a gene dosage imbalance between male and female organisms at X-linked loci. Therefore, it is likely that dosage compensation mechanisms have evolved in parallel with hetereomorphic sex chromosomes in order to equalize gene dosages between the sexes.

In mammalian embryos dosage compensation is achieved by a process known as X chromosome inactivation (XCI). As stated in the Lyon hypothesis one of the two female X chromosomes is transcriptionally inactivated, leaving a single active X for both male and female individuals (Lyon 1961). Due to its significance XCI is tightly regulated. Interestingly, there is a lot of variation in the molecular mechanisms of XCI between the different orders of the eutherian clade. To date, the house mouse (*Mus musculus*, short *mm*) is one of the best studied model organisms for this process (Payer and Lee 2008). Therefore, this thesis work focusses on the murine XCI mechanism.

1.1.3 Escape from X inactivation

In mice XCI is the reason for an individual's relative tolerance to the loss of one of the female X chromosomes (Lyon and Hawker 1973, Burgoyne and Baker 1981, Burgoyne, Evans et al. 1983). Consequently, XO mice are viable and fertile even though the number of germ cells is reduced by approximately 50%. In humans, however, the XO condition (Turner syndrome) is almost always lethal, and over 99% of conceptions are aborted in the first trimester of pregnancy (Hunt 1991). This discrepancy between the two species may be explained by the

different rates of X inactivation escape in the two species. Embryonic lethality and Turner's syndrome are thought to be due to monosomy of gene(s) common to the X and Y chromosome (Ferguson-Smith 1965). In healthy XX females these genes are expressed from both alleles. *ZFX* and *RPS4*X are genes with this property that play a role in Turner's syndrome (Burgoyne 1989, Fisher, Beer-Romero et al. 1990). In contrast, both *Zfx* and *Rps4X* undergo normal XCI in mice (Ashworth, Rastan et al. 1991). Whereas human *ZFX* and *RPS4X* possess Y-linked homologues, Y-linked genes are absent in mouse. Therefore, humans achieve dosage compensation between males and females by escape of the X-linked homologues from XCI. Consequently, XO humans suffer from haploinsufficiency for the two genes probably contributing to the observed low viability in affected individuals. XO mice on the other hand are not different to XX or XY individuals with respect to *Zfx* and *Rps4X* expression because of the monoalellic expression of these genes (Ashworth, Rastan et al. 1991).

In man a number of loci are known to evade XCI. These fall into three main categories: 1) Firstly, there are loci with a Y-linked homologue within a homologous region, i.e. a pseudoautosomal locus. 2) Secondly, there are loci that have a Y-linked copy in a non-homologous region. 3) And lastly, there are human escapees without a Y-linked copy.

1) Regions of homology between X and Y are pseudoautosomal regions (Ellis and Goodfellow 1989) where gene dosage is balanced between the sexes. In fact, pseudoautosomal genes like *MIC2* do not undergo XCI in humans (Goodfellow, Pym et al. 1984).
2) Examples for genes in this second class are *ZFX* and *ZFY* and *RPS4X* and *RPS4Y* (Page, Mosher et al. 1987, Palmer, Sinclair et al. 1989).
3) This group consists of genes that are expressed at levels that are twice as high in females as compared to males; an example is *UBE1 (Zacksenhaus and Sheinin 1990)*.

Most of the mouse homologues of the human escapees, however, undergo faithful inactivation (Ashworth, Rastan et al. 1991, Zinn, Bressler et al. 1991). Nevertheless, mouse XCI is incomplete as well. More recently, the proportions of escaping genes in the two species have been determined: Whereas about 15% of human X-linked genes escape silencing, only 3% escape in mice (Carrel and Willard 2005, Berletch, Yang et al. 2011). The difference in XCI escape between human and mouse may explain the observed phenotypical differences in XO human and mouse females.

1.1.4 X chromosome inactivation during mouse embryonic development

There are many studies that provide insight into the dynamics of XCI in the mouse embryo. Amongst others, it was found that in mice there are two modes of XCI: In one of which the future Xi is selected by chance (random) and in the other, the origin of the Xi is pre-determined (imprinted) (Figure 1.1) (reviewed in Huynh and Lee 2001). Both random and imprinted XCI are spatially separated, i.e. they happen in different tissues. Whereas random XCI occurs in somatic cells of the embryo proper, imprinted XCI takes place in the pre-implantation embryo and the extraembryonic tissue (trophectoderm and primitive endoderm) (Figure 1.1). In contrast to random XCI, where either the X of maternal (Xm) or the X of paternal (Xp) origin is inactivated, it is always the Xp that is silenced in imprinted XCI.

Imprinted XCI occurs before random XCI as it is initiated in the pre-implantation embryo (reviewed in Thorvaldsen, Verona et al. 2006). To date, two models have been proposed to describe the mechanism of imprinted XCI. In the first model the Xp is silenced through meiotic sex chromosome inactivation (MSCI) (reviewed in Handel 2004). Therefore, the gametic Xp is in a pre-inactivated state already, and the corresponding model is called pre-inactivation hypothesis (reviewed by Huynh and Lee 2001). The major drawback of the pre-inactivation hypothesis is that in the early embryo both X chromosomes are active until the mid-cleavage stage (Monk and Harper 1978). This means that, if this model was true, an

additional re-activation step right after fertilization would be required. The second model is the *de novo* inactivation hypothesis. It promotes the idea that the silent state of the Xp is acquired upon fertilization (Okamoto, Otte et al. 2004). In this case an extra re-activation step would not be required. Thus, working on the principle of parsimony *de novo* inactivation deems to be the more likely mode for imprinted XCI.

Upon implantation of the mouse embryo, random XCI is initiated. It is a complex, multi-step process that reaches completion at 5.5 to 6.5 *days post coitum* (*dpc*) (Rastan 1982, Takagi, Sugawara et al. 1982). Random XCI starts off with a counting step ensuring that a single X chromosome remains active. This is followed by a choice-step at which cellular mechanisms decide whether the Xp or the Xm will be silenced.

Since, embryonic stem cells (ESCs) are derived from the inner cell mass (ICM), where cells undergo random XCI, they are a widely used model system for the investigation of this process. ESCs are an especially useful tool because their random XCI can be switched on through differentiation. That is why many studies utilize mouse ESCs to investigate the molecular mechanism of XCI *in vivo*.

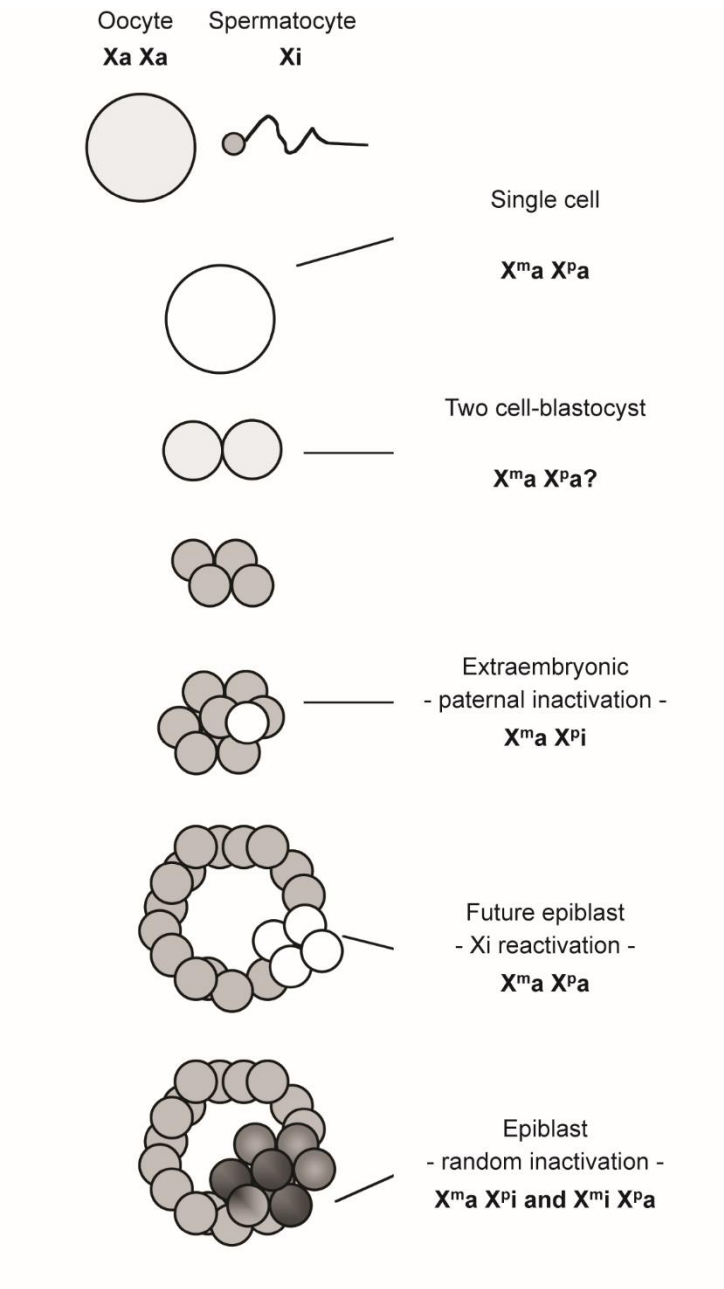


Figure 1.1 – XCI during mouse embryonic development.

Schematics showing the fate of maternal (Xm) and paternal (Xp) X chromosomes during early mouse development starting from egg and sperm. Note, that initially the oocyte carries two active X chromosomes (Xa Xa) and the sperm an inactivated chromosome (Xi). Right after syngamy the zygote bears two active X chromosomes, one of which is of paternal and the other of maternal origin. During subsequent cell divisions daughter cells become committed to either the extraembryonic (imprinted XCI) or epiblast (random XCI) lineages (Figure adapted from Chow et al. (Chow, Yen et al. 2005)).

1.2 The inactive X chromosome and Xist RNA

By now it has become clear that the Xi is a special chromosome. In order to gain a better understanding of what makes the Xi unique, it is worthwhile to also look into the basics of Chromatin Biology.

1.2.1 Higher-order chromatin structure at the female X chromosomes

In the 19th century cell biologists discovered an easily stainable substance in the cell nucleus. Owing to this property Walther Flemming named the structure chromatin (reviewed in Trojer and Reinberg 2007). In terms of Biochemistry, chromatin is a complex assembly of macromolecules whose basic subunits are the nucleosomes. The nucleosome is an octameric protein complex consisting of dimers of the four types of core histones; histone 2A (H2A), histone 2B (H2B), histone 3 (H3) and histone 4 (H4) around which a stretch of 147 base pair (bp)-long deoxyribonucleic acid (DNA) is wound (Figure 1.2A). Hereby, H2A and H2B form heterodimers, whereas H3 and H4 have been found to homodimerize. Individual nucleosomes are interconnected by a stretch of linker DNA (Luger, Mader et al. 1997, Davey, Sargent et al. 2002). An array of nucleosomes is believed to fold into a highly compacted structure, the 30 nm fiber (Figure 1.2C) (Finch and Klug 1976, Thoma, Koller et al. 1979). This fiber has the potential to form higher-order structures that eventually make up the chromosomes.

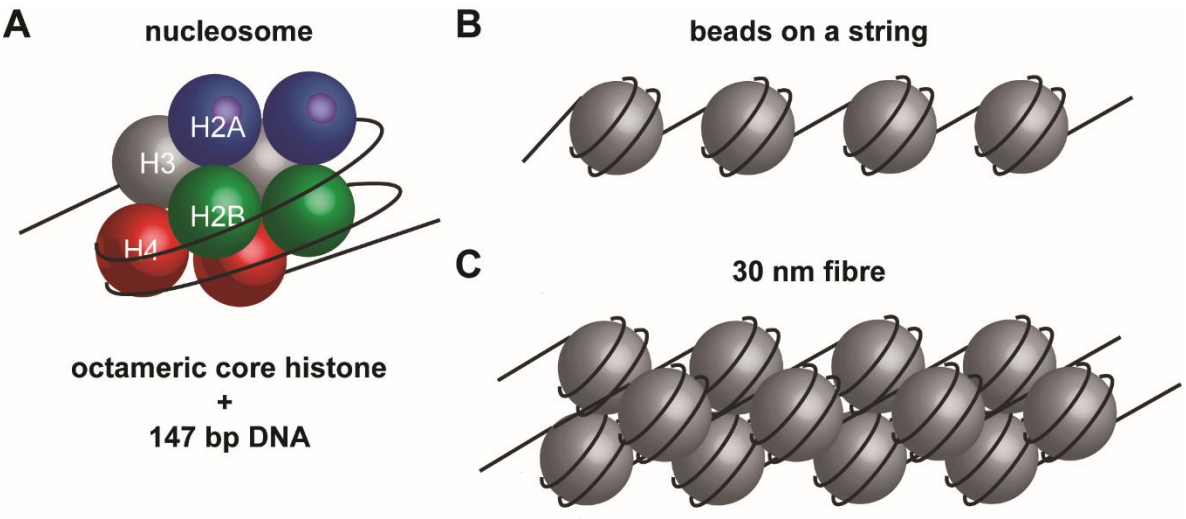


Figure 1.2 - Nucleosome structure and chromatin.

(A) Illustration of a nucleosome consisting of the core histones H2A (blue), H2B (green), H3 (grey) and H4 (red). An about 147 base pair (bp) -long stretch of DNA is wrapped around the octameric structure. (B) Individual nucleosomes are linked together by a piece of linker DNA, and together they form a "beads-on-a-string" conformation. (C) The "beads-on-a-string" are further folded together into a structure that is about 30 nm in diameter. This 30 nm fiber is the building block for even more complex chromatin folds (reviewed in Jansen and Verstrepen 2011).

Chromatin folding and compaction are complex processes that lead to different levels of chromatin compaction. Nowadays, a distinction is made between various types of chromatin: On the one hand there is the relatively decompacted euchromatin (EC), and on the other hand there is the strongly condensed heterochromatin (HC) (reviewed in Trojer and Reinberg 2007).

EC was found to be open and easily accessible to the transcriptional machinery (reviewed in Trojer and Reinberg 2007). HC, by contrast, is characterized through a more closed structure that is rather inaccessible for factors involved in transcription. It is, however, important to note that there are many other intermediate states of compaction, and exemplifying chromatin as existing in two states is a simplification. For this reason researchers came to distinguish constitutive and facultative HC (cHC and fHC). Hereby, cHC is the permanently silent part of the chromatin. In higher eukaryotes this includes telomeres, centromeres, repetitive and non-coding elements (Kent, Sugnet et al. 2002, reviewed in Craig 2005, Martens, O'Sullivan et al. 2005, reviewed in Talbert and Henikoff 2006). Even though cHC is very condensed, some basic transcriptional activity has been observed in these regions. FHC, on the contrary, retains the ability to convert into transcriptionally active EC.

The Xi is one of the classical examples for fHC. The DNA in the chromatin of this chromosome exhibits increased levels of DNA methylation, an epigenetic mark generally associated with transcriptional silencing, at CpG islands and promoters (reviewed in Trojer and Reinberg 2007). Furthermore, the chromosome bears an array of repressive histone modifications, for example histone 3 lysine 27 trimethylation (H3K27me3) and histone 2A lysine 119 monoubiquitination (H2AK119ub1). There are also histone 3 lysine 9 dimethyl (H3K9me2) and histone 4 lysine 20 monomethyl (H4K20me1) marks. Additionally, a global depletion of acetylation marks (hypoacetylation) and H3K4me3, located to active promoters, at the core histones of the Xi have been observed.

In most cases DNA and chromatin signatures cannot alter the chromatin structure by themselves. *Ergo*, there must be other factors that translate the histone code into biological information (reviewed in Rothbart and Strahl 2014). This is when so called reader proteins come into play. These are factors that comprise domains that specifically bind to histone marks following a lock-and-key principle (reviewed in Taverna, Li et al. 2007). Oftentimes, the binding of histone code readers results in the recruitment of writers, which are proteins bearing domains capable of setting certain histone marks. Prominent examples for a family of readers and writers that play a role in the silencing of the Xi are the Polycomb Group (PcG) Proteins that are discussed in more detail in Section 1.2.3.

Together PcG proteins form the Polycomb repressor complexes 2 and 1 (PRC2 and 1) that mediate gene silencing. Both complexes are active as writers, setting the silencing H3K27me3 and H2AK119ub1 marks, respectively. Briefly, the PRC2 core consists of different isoforms of Embryonic ectoderm development (EED), of Enhancer of zeste 1 or 2 (Ezh1/2), Suppressor of zeste 12 (Suz12) and of either Retinoblastoma binding protein 4 or 7 (RBBP4/7, also known as RbAp48/46) (Wang, Wang et al. 2004). The catalytic core of PRC1 is formed by the Really interesting gene 1 (Ring1, RING1A or RING1B in mammals) proteins and Posterior sex combs (Psc, PcG Ring Finger [PCGF] homologues in mammals) (Wang, Wang et al. 2004). The Ezh1/2 proteins are at the heart of PRC2 activity. The two

proteins are responsible for setting H3K27me2 and H3K27me3 marks (Cao, Wang et al. 2002). Ezh1/2 activity is contained in the Su(var)3-9, Enhancer-of-zeste, Trithorax (SET) domain. Interestingly, the PRC2 component EED can bind to H3K27me3 marks through its WD40 beta propeller repeats, an interactions that had been proposed to stimulate PRC2's methyltransferase activity (Margueron, Justin et al. 2009, Xu, Bian et al. 2010).
Similarly, the PRC1 component RING1B places H2AK119ub1 marks at places where PRC2 had marked the chromatin with H3K27me3 (Wang, Wang et al. 2004, Cao, Tsukada et al. 2005, Zou, Cao et al. 2005). These are believed to be bound by the mammalian, chromodomain-containing Psc homologues (Fischle, Wang et al. 2003, Min, Zhang et al. 2003).

Unmethylated CpG sequences can act as recruitment signal for PRC2 (Bartke, Vermeulen et al. 2010, Lynch, Smith et al. 2012). The Xi exhibits a high degree of DNA methylation at CpG islands and promoters (Trojer and Reinberg 2007). In light of the finding that PRC2 factors are enriched at the Xi, unmethylated CpG islands cannot be the only docking sites for the silencing complex.

The mechanisms of maintaining and propagating silencing signatures are important for the formation of a silencing compartment at the Xi. Therefore, it is of utmost importance that the epigenetic code is set and read correctly (Chaumeil, Le Baccon et al. 2006, Clemson, Hall et al. 2006).

1.2.2 The X inactivation center and the discovery of Xist RNA

The concept of a *cis*-acting X inactivation center (Xic in mouse, XIC in man) as chromosomal region from which XCI is initiated was proposed by Russell in 1963 (Russell 1963). He based his hypothesis on X;autosome translocation studies where inactivation was seen to spread from X-linked material into adjoined autosomal regions, thus, inactivating normally active autosomal loci. Importantly, the silenced translocation product must carry the Xic (Russell 1963). Later Lyon *et al.* (1986) demonstrated the *cis*-limited action for a true X-

linked locus in genetic studies (Lyon, Zenthon et al. 1986). Their study looked at alleles of the *sparse fur* (*spf*) locus in conjunction with the T(X;4)37H and T(X;11)38H translocations that have breakpoints on either side of *spf*. Lyon *et al.*'s work was underpinned by cytogenetic studies on X;autosome translocations using heterochromatinization and late replication as indicators for XCI (Russell and Cacheiro 1978, Disteche, Eicher et al. 1979, Rastan 1983).

In the original publications the positions of both mouse and human Xic/XIC had been determined genetically. By looking at translocations Rastan (1983), and Rastan and Robertson (1985) mapped the Xic to an approximately 10 centimorgan (cM) region on the X chromosome (Rastan 1983, Rastan and Robertson 1985). Later this was refined to roughly 8 cM (Keer, Hamvas et al. 1990, Rastan and Brown 1990). More recently the murine Xic sequence was narrowed down to a 100 to 500 kb domain (Chureau, Prissette et al. 2002). The human XIC, on the other hand, had been delineated in studies on X chromosomal rearrangements. These identified a 0.5 Mb region on the human X that is syntenic to the locus previously identified in mouse (Brown, Lafreniere et al. 1991, Leppig, Brown et al. 1993).

In addition to identifying an X-linked chromosome section responsible for XCI, it was also proposed that within this region there is a component that exerts control over the entire mechanism. Upon characterization, this region was termed *X controlling element* (*Xce*) (Cattanach and Isaacson 1967, Cattanach, Pollard et al. 1969, Cattanach, Perez et al. 1970). Later serendipity led researchers to the discovery of a long-noncoding ribonucleic acid (lncRNA) within the human *Xce*. This lncRNA is X-linked and, importantly, expressed exclusively from the Xi (Brown, Ballabio et al. 1991). Subsequently the large, nuclear, apparently Xi-associated transcript that lacked protein-coding potential was also identified in mice and termed X inactive specific transcript (XIST in human and Xist in mouse) (Brockdorff, Ashworth et al. 1991). Closer investigations showed that full-length human XIST

RNA is 17 kb long (Brown, Hendrich et al. 1992). With a length of 15 kb murine Xist is shorter than its human homologue (Figure 1.3) (Brockdorff, Ashworth et al. 1992). Human and mouse XIST/Xist RNA contain eight and seven different short tandem repeat domains, respectively. The majority of these domains play a redundant role in XIST/Xist RNA localization along the Xi (Wutz, Rasmussen et al. 2002). But the exact mechanisms of XIST/Xist association with the chromosome elect have not been elucidated. The best studied tandem repeat domain is the A-repeat (RepA) in exon 1 that is required for Xist-mediated chromosome silencing but not for Xist localization (Wutz, Rasmussen et al. 2002).

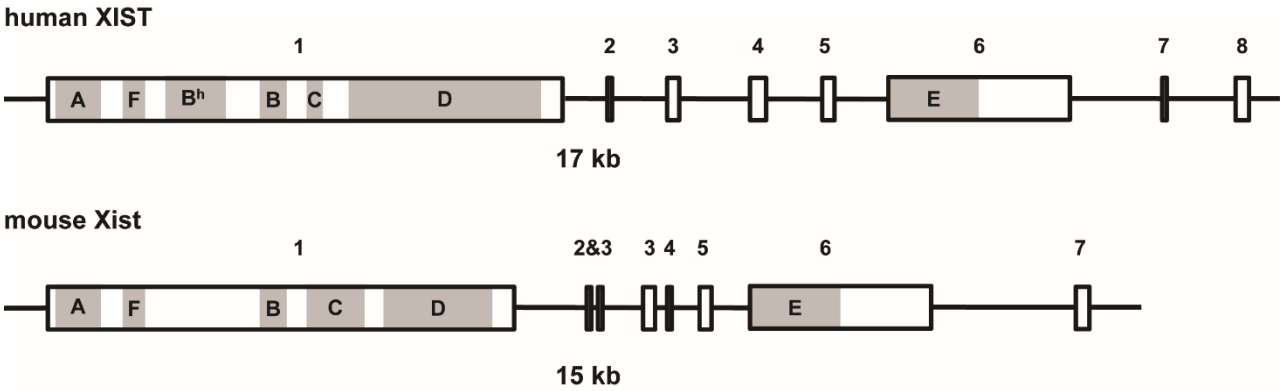


Figure 1.3 – Human and mouse XIST/Xist RNA gene structures.

Comparison between gene structures of human and mouse XIST/Xist RNA. Boxes represent exons (numbered), introns are represented by lines and repeat elements are displayed in grey boxes (letters A through F). Schematics are not drawn to scale (reviewed in Brockdorff 2002).

Upon Xist RNA expression and spreading *in cis* along the X chromosome elect, the chromosome territory is depleted of RNA polymerase II, splicing factors and active chromatin marks, such as histone acetylation, are removed as well (Jeppesen and Turner 1993, Clemson, McNeil et al. 1996, Heard, Rougeulle et al. 2001, Kucera, Reddy et al. 2011). To date, the underlying molecular mechanisms of this process have remained elusive. Moreover, it has been suggested that the absence of transcription is sufficient to initiate chromatin compaction because the size of the Xi had been reduced in actinomycin D-treated

cells (Naughton, Sproul et al. 2010). Subsequently, large multi-protein silencing complexes are recruited. The best studies amongst those are the PRC1 and PRC2. As discussed earlier, PRC1/2 set the prominent silencing marks H3K27me3 and H4K119ub1 that are part of a complex, down-stream silencing cascade (Mak, Baxter et al. 2002, Plath, Fang et al. 2003, Silva, Mak et al. 2003, de Napoles, Mermoud et al. 2004).

Once established, the Xi-state is maintained and propagated through many rounds of cell division. Because faithful XCI is vital for the cell, the entire process is tightly regulated, and Xist RNA cannot induce XCI in already differentiated cells (Wutz and Jaenisch 2000, Kohlmaier, Savarese et al. 2004). In the literature the time frame during which the chromosome is susceptible to Xist-mediated silencing is often referred to as “window of opportunity”. Wutz and Jaenisch (2000) showed that in a transgenic mouse embryonic stem cell (ESC) system Xist RNA cannot induce transcriptional repression after 48 hours of differentiation, and after 72 hours it is also not required anymore for the maintenance of the silent state, making Xist RNA a factor that is primarily involved in the early initiation and spreading stages of XCI (Wutz and Jaenisch 2000). After completion of these initial stages other factors may be involved in the maintenance of the inactive state.

1.2.3 Polycomb group proteins in X chromosome inactivation

During the maintenance phase of XCI the Xi starts to replicate late and accumulates histone variant macroH2A (Costanzi and Pehrson 1998, Csankovszki, Panning et al. 1999, Mermoud, Costanzi et al. 1999). Important roles in the maintenance phase have been assigned to the aforementioned PcG complexes. Because it was found that Xist expression is required for PcG protein recruitment, it has been proposed that the lncRNA interacts with these proteins directly and, thus, recruits them to the Xi (Plath, Fang et al. 2003, Silva, Mak et al. 2003, Schoeftner, Sengupta et al. 2006, Zhao, Sun et al. 2008, Hasegawa, Brockdorff et al. 2010, Kaneko, Li et al. 2010). Evidence for the direct recruitment of PcG proteins through the RepA of Xist RNA came from *in vitro* experiments with individual PRC2 subunits

and from *in vivo* crosslinking studies (Zhao, Sun et al. 2008, Kaneko, Li et al. 2010, Kanhere, Viiri et al. 2010). However, especially the *in vitro* experiments presented in the current literature are prone to artifacts, e.g. through non-physiological complex formation in co-immunoprecipitation (co-IP) experiments. Thence, it is difficult to know whether the identified interactions are truly direct. Furthermore, Xist RNA with a RepA deletion is able to recruit PcG proteins at reduced levels (Plath, Fang et al. 2003, Kohlmaier, Savarese et al. 2004, Schoeftner, Sengupta et al. 2006, Zhao, Sun et al. 2008). This observation would argue for an intermediate interaction between the lncRNA and the repressive protein complex. Although the X was not properly silenced when Xist RNA lacking RepA is expressed, a depletion of RNA polymerase II and a change in epigenetic signatures at the chromosome were observed (Chaumeil, Le Baccon et al. 2006, Pullirsch, Hartel et al. 2010). These changes at the X chromosome may account for the observed PcG recruitment, leaving open the question of whether Xist RNA is a direct PRC1/2 recruiter or not.

1.2.4 Protein factors involved in X chromosome inactivation

Aside from PcG proteins, not many other potentially Xist/XIST-interacting and XCI maintenance factors have been identified. For years, researchers have been looking for such candidates using empirical and *in silico* approaches (Pullirsch, Hartel et al. 2010, Chan, Zhang et al. 2011, Agostini, Cirillo et al. 2013). Thereby prominent Xi-enrichment has been one of the most commonly used initial determinants for relating a candidate protein with XCI (Blewitt, Gendrel et al. 2008, Pullirsch, Hartel et al. 2010, Chan, Zhang et al. 2011). Amongst others, Chan *et al.* (2011) have conducted a genome-wide RNA interference (RNAi) screen revealing a set of 32 proteins with putative roles in the maintenance of XCI (Chan, Zhang et al. 2011). A majority of these 32 candidate proteins was involved in either cell cycle regulation, gene transcription, chromatin related functions or RNA processing. However, none of them seemed to directly interact with Xist RNA. Royce-Tolland *et al.* (2010), on the contrary, demonstrated that the Alternative Splicing Factor/Splicing Factor-2 (ASF/SF2) interacts with the conserved RepA of Xist (Royce-Tolland, Andersen et al. 2010).

Interestingly, this is the same repeat region that may be recognized by the PRC2 component Ezh2 (Zhao, Sun et al. 2008, Maenner, Blaud et al. 2010). Whereas Ezh2 binding affinity was reported to be RNA secondary structure-dependent, the ASF2/SF2-Xist RNA associations seem to rely on primary sequence recognition (Royce-Tolland, Andersen et al. 2010). Importantly, these findings suggest that the A-repeat is vital for Xist RNA function. Furthermore, the association with a splicing factor shows that faithful post-translational processing is paramount in the Xist RNA-dependent XCI process.

The Scaffold Attachment Factor-A/heterogeneous ribonucleoprotein U (SAF-A/hnRNP U) associates with the Xi and potentially binds directly to Xist RNA (Hasegawa, Brockdorff et al. 2010). Previously, it had been demonstrated that SAF-A/hnRNP U is involved in splicing processes (Xiao, Tang et al. 2012). This finding may relate to the need for Xist RNA to be processed accurately on a post-translational level. SAF-A/hnRNP U is a versatile protein. It has not only been linked to splicing, but is also believed to be a major structural component of the nucleus (for a more detailed discussion of SAF-A/hnRNP U function c.f. Chapter 2) (Dreyfuss, Matunis et al. 1993). Importantly, SAF-A/hnRNP U comprises a DNA and a RNA binding domain. The presence of these domains suggests that the SAF-A/hnRNP U can bind to DNA and RNA concomitantly in order to fulfill its role as a molecular scaffold within the nucleus. In respect to a potential Xist RNA association, this implies that SAF-A/hnRNP U binds to Xist *via* its RNA binding domain and then tethers the lncRNA to chromatin through the DNA binding domain (Hasegawa, Brockdorff et al. 2010). A similar mode of action has been proposed for the Ying Yang 1 (YY1) protein (Jeon and Lee 2011). The transcriptional repressor YY1 has a DNA binding domain, which can also bind RNA, and through which it is believed to attach Xist RNA to its specific binding sites on the Xi.

Apart from PcG proteins, it has also been reported that Absent, small or homeotic discs 2-like (Ash2l) localizes to the human Xi at the maintenance phase of XCI (Pullirsch, Hartel et al. 2010). Ash2l is a member of the Trithorax Group protein family (TrxG) that antagonizes

PcG proteins (reviewed in Schuettengruber, Chourrout et al. 2007). Like PcG complex proteins, Ash2l can also be recruited by mutated Xist RNA that does not induce gene silencing (Pullirsch, Hartel et al. 2010). The concurrence of PcG complexes and a TrxG protein is surprising. However, to date Ash2l is the only identified Xi-enriched TrxG protein, thus, leaving open the possibility that Ash2l is being recruited independently.

Another protein that is not directly linked to Xist, but was shown to play some role in XCI, is Structural maintenance of chromosomes flexible Hinge Domain containing 1 (Smchd1) (Blewitt, Gendrel et al. 2008). It had been reported that SmcHD1 is required for the maintenance of CpG island DNA methylation and gene repression at the Xi. However, the protein's exact molecular role in XCI has not been revealed, yet.

Lastly, there is the Special AT-rich Binding 1 protein (SATB1) that influences Xist RNA localization patterns in the nuclei of lymphocytes (Agrelo, Souabni et al. 2009, Arthold, Kurowski et al. 2011). In this particular case, loss of SATB1 in tumor cells abrogates the silencing function of Xist. However, there is limited evidence to whether SATB1 function extends to other cellular systems.

Recently, Agostini *et al.* (2013) developed an algorithm to predict Xist RNA-protein interaction motifs (Agostini, Cirillo et al. 2013). Amongst others they showed that the PRC2 components Suz12 and Ezh2 have strong propensities to bind RepA of Xist RNA, thus, supporting the view that PRC2 and Xist interact directly (Maenner, Blaud et al. 2010). Likewise, the ASF/SF2-Xist interaction with the RepA region was confirmed by the algorithm (Royce-Tolland, Andersen et al. 2010, Agostini, Cirillo et al. 2013). YY1, on the contrary, showed a higher interaction propensity for repeat region C. Albeit SAF-A/hnRNP U may act in a mode similar to YY1, their predicted Xist RNA binding sites differ. The computational analysis showed that the interaction fragments identified by Hasegawa *et al.* (2010) matched the computationally determined Xist regions that are most prone to interact with SAF-A/hnRNP U (Hasegawa, Brockdorff et al. 2010, Agostini, Cirillo et al. 2013). Interestingly,

Agostini *et al.* also suggest an interaction of SAF-A/hnRNP U with the RepA loops. These repeat loops appear to be the main anchor points for Xist-interacting proteins as also SATB1 has been suggested to bind there (Agostini, Cirillo et al. 2013).

In the past, pieces of evidence have been collected that either relate particular proteins to a certain stage of XCI or even suggest an immediate interaction between these polypeptides and the lncRNA. Experimental evidence and computational predictions suggested that there are multiple binding sites within the Xist RNA sequence for a given protein (Hasegawa, Brockdorff et al. 2010, Agostini, Cirillo et al. 2013). This observation leads to the conclusion that several binding sites may be involved in cooperative interactions, eventually mediating chromosome-wide silencing.

1.2.5 Xist RNA spreading and LINE repeats

The inactivation signal in mouse X;autosome translocations was seen to spread from X chromosomal into attached autosomal regions (Russell 1963). However, autosomal regions further away from the X;autosome breakpoint tend to be less influenced by the inactivation process than those that are closer (Cattanach 1961, Russell 1963). These observations led Gartler and Riggs (1983) to put forward the idea of “booster” elements or “way stations” for the propagation of X inactivation along the X chromosome (Figure 1.4) (reviewed in Gartler and Riggs 1983, Riggs, Singer-Sam et al. 1985). They suggested that these were conserved elements, concentrated around the Xic and along the entire X chromosome that were absent from autosomes.

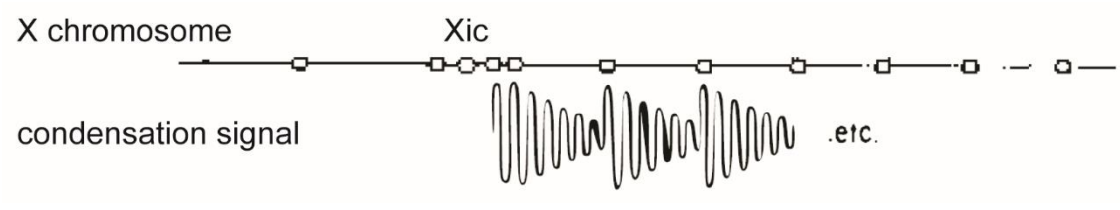


Figure 1.4 – Gartler and Riggs model for the propagation of XCI.

A modified schematics from the original publication is shown where the X chromosome is displayed as a line with an enrichment of XCI "booster" elements (white boxes). X chromosome condensation is initiated from the X inactivation center (Xic) and propagated from one "booster" element to the other. The signal strength is displayed as a wave function with changing amplitude (reviewed in Gartler and Riggs 1983).

Later, the Gartler-Riggs model was refined by Mary Lyon who proposed long-interspersed elements 1 (LINE1) repeats as candidates for the "booster" elements/"way stations" (reviewed in Lyon 1998). LINE1 elements are autonomous, non-long terminal repeat (LTR) retrotransposons of about 6 to 7 kb length, some of which have retained the ability to amplify themselves in the host genome (Burton, Loeb et al. 1986). Lyon's "repeat hypothesis" was based on two observations:

1) FISH imaging studies on mouse and human cells using LINE1 probes showed that the X chromosomes of these cells were stained more intensely than any of the autosomes (Korenberg and Rykowski 1988, Boyle, Ballard et al. 1990). Therefore it was concluded that LINE1 elements were more enriched on the X chromosomes as compared to the autosomes.
2) A re-examination of X:autosome translocation data lead Mary Lyon to the conclusion that X inactivation cannot propagate onto autosomes when LINE1 repeats are missing.

In fact, a later study showed that LINE1s are distributed non-randomly on the X chromosome, and in humans a near two-fold enrichment of this class of repeat elements had

been observed at the X compared with other autosomes (Bailey, Carrel et al. 2000). In line with Gartler and Riggs' theory an especially high density of such elements was observed around the Xic. Furthermore, loci that are known to escape XCI were depleted of LINE1 elements. Later on, Chow *et al.* (2010) showed that one of the initial stepping stones in heterochromatin formation at the Xi is the appearance of silent LINE and short interspersed element (SINE) repeats within the Xist-covered regions of the chromosome (Chow, Ciaudo et al. 2010). This re-localization event was seen prior to gene inactivation, and is in line with the observation that genic regions on the X are initially localized to the outside of the Xist RNA territory, but move into the territory upon inactivation (Chaumeil, Le Baccon et al. 2006). Thus, a high LINE density may provide a permissive environment for heterochromatinization, and silencing compartment formation. Chow *et al.* (2010) added another level of complexity by discovering a subset of LINEs that is specifically expressed from the Xi (Chow, Ciaudo et al. 2010). These young, active LINEs facilitate the silencing of certain genes, possibly through an RNA interference (RNAi) mechanism. A refined map of Xist localization was prepared by Simon *et al.* (2013) who used capture hybridization analysis of RNA targets with deep sequencing (CHART-seq) to map genome-wide binding sites of Xist RNA (Simon, Pinter et al. 2013). With oligonucleotides complementary to Xist RNA they captured RNA-associated DNA at four distinct developmental stages. Apart from providing insight into Xist RNA localization, these experiments also identified the region of Xist predominantly associated with target sites: It was exon 1 that was most enriched in CHART. In the course of differentiation, Xist was found to first target gene-rich regions, only then gene-poor domains were covered. Furthermore, it seems that gene bodies of XCI escapees are spared by Xist, with the RNA localizing around the borders of these genes (Simon, Pinter et al. 2013). A similar study was conducted by Engreitz *et al.* (2013) who utilized RNA antisense purification (RAP). Like CHART, RAP uses complementary oligonucleotide probes to capture the lncRNA and associated DNA sequences (Engreitz, Pandya-Jones et al. 2013). They explored Xist spreading behavior during initiation and maintenance phases of XCI. By RAP it was demonstrated that somatic cell Xist was

enriched across the entire X with local differences in enrichment. Especially enriched sites correlated with H3K27me3 marks and gene density. However, there was a negative correlation between Xist and LINE sequences. Moreover, escapees were consistently Xist-depleted with approximately 50% less occupancy than on silent X-linked genes. Analysis of initial Xist spreading by RAP showed that over time Xist spread over the entire chromosome until a maintenance-like pattern was achieved. Eventually, Engreitz *et al.* used their RAP data to address the question of whether Xist spreads linearly along a certain set of high-affinity sites ("affinity transfer") or moves from one site to the next proximal point of attachment, based on chromosome conformation ("proximity transfer"). Their analysis could not provide evidence for an "affinity transfer"-like mechanism. But the "proximity transfer" model was substantiated by a combined evaluation of previously obtained Hi-C chromosome conformation capture and RAP data (Lieberman-Aiden, van Berkum et al. 2009). The resulting model of Xist propagation suggests that Xist exploits the three-dimensional arrangement of chromosome loops to facilitate initial spreading. Later Xist-binding factors come in facilitating the re-arrangement of the entire chromosome and rendering it inactive through a more compact conformation (Engreitz, Pandya-Jones et al. 2013).

1.2.6 Xist RNA localization during the cell cycle

Upon differentiation, Xist RNA starts spreading *in cis* along the selected X chromosome. Once spreading is complete, Xist remains associated with the Xi throughout interphase (Figure 1.5) (Clemson, McNeil et al. 1996).

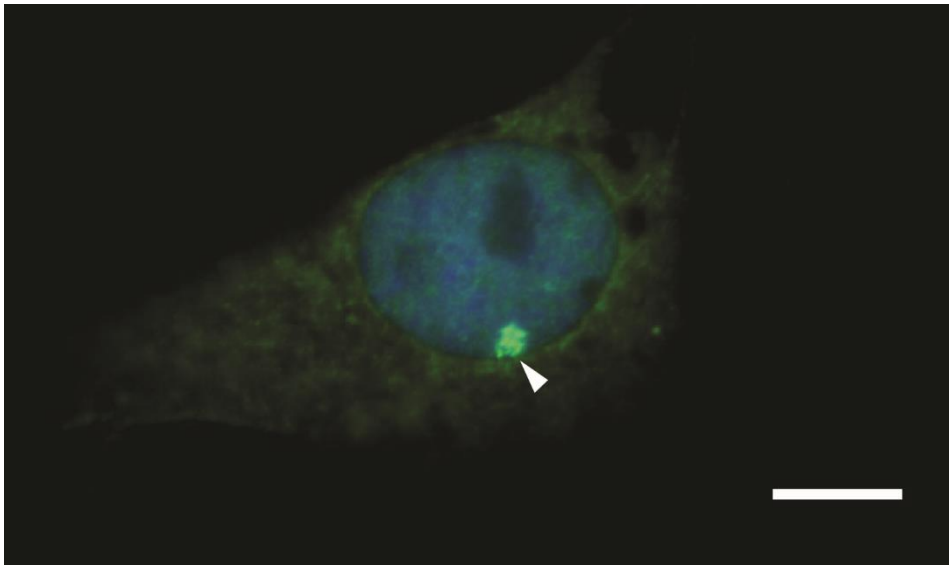


Figure 1.5 – Xist RNA interphase localization.

Interphase, mouse fibroblast with Xist RNA FISH (green) and 4', 6-diamidino-2-phenylindole (DAPI)-stained DNA (blue). The arrowhead (▼) points at Xist RNA cloud.

In 1996 it was shown that Xist RNA localization and stable association with the interphase Xi are facilitated by the nuclear matrix. The nuclear matrix, often referred to as nuclear scaffold, is a proteinaceous structure that remains after treating cell nuclei with nucleases and high-salt buffers (for a more detailed discussion c.f. Section 1.3) (Berezney 1991). Also characteristic of the interphase Xi, and therefore Xist interphase positioning, is its localization to perinuclear regions during the gap 1 phase (G1 phase) of the cell cycle. This changes at the beginning of the synthesis phase (S phase), when the Xi moves to a perinucleolar location (Zhang, Huynh et al. 2007). The observed change in sub-nuclear positioning hints at that the Xi interacts with structural components of the nucleus, possibly through an interaction with the nuclear matrix in conjunction with Xist RNA. This would also explain why a deletion of Xist results in an aberrant localization of the Xi (Zhang, Huynh et al. 2007). Even more interesting is that autosomes expressing a transgenic *Xic* also localize to perinucleolar regions (Lou, Li et al. 2007).

When a cell enters the division phase (mitosis) chromosome shape and condensation change (reviewed in Civelekoglu-Scholey and Cimini 2014). But what happens to Xist RNA? - Although the lncRNA is confined to the Xi, which is propagated to the daughter cells, several studies have indicated that Xist RNA dissociates from the Xi during mitosis. Immuno-fluorescence *in situ* hybridization (FISH) experiments revealed that human XIST RNA dissociates from the Xi at prophase (Hall, Byron et al. 2009). Rodent Xist, on the contrary, was shown to remain close to the Xi until anaphase (Duthie, Nesterova et al. 1999). Eventually, at telophase, also rodent Xist has completely dissociated from the Xi. At this point Xist molecules are distributed in a particulate pattern throughout the nucleus (Duthie, Nesterova et al. 1999, Hall and Lawrence 2003, Smith, Byron et al. 2004). These findings imply that in human and mouse cells XIST/Xist RNA is stripped off the Xi for a short time during mitosis and freely diffuses through the cell. Jonkers *et al.* (2008), however, published an opposing viewpoint based on studies in differentiated mouse ESCs (Jonkers, Monkhorst et al. 2008). In their experimental setup Xist associated with the initial Xi right from the start of mitosis and remained there until the end. Furthermore, Xist RNA seemed to be transcribed at a constant rate throughout all cell cycle stages. Most likely, different methodologies lead to the two contrasting outcomes (Duthie, Nesterova et al. 1999, Jonkers, Monkhorst et al. 2008). Nevertheless, also Jonkers *et al.* agree on that binding of Xist to the Xi is less tight at the end of mitosis (Heath, Ribeiro de Almeida et al. 2008).

1.3 The nuclear matrix/scaffold – foundation for nuclear organization

During the past 30 years the conception of the internal organization of an interphase nucleus has changed from a random arrangement of chromatin embedded in the nucleoplasm to a chromatin loop domain model, where chromatin is organized into loops that are anchored in a structure named nuclear scaffold or matrix (Gasser and Laemmli 1986, Berezney 1991). The nuclear scaffold/matrix was first isolated in 1974 by Berezney and Coffey, who defined the matrix as fibrous, proteinaceous network that remains after treating cell nuclei with high-salt buffers, detergent(s) and nuclease(s) (Figure 1.6) (Berezney and Coffey 1974, Berezney

and Coffey 1977, Mirkovitch, Mirault et al. 1984). The treatment removes nearly all chromatin, DNA, RNA and phospholipid components from cell nuclei and leaves a structure that is almost solely made of protein. Due to the non-physiological nature of the procedure, the functional relevance of the nuclear scaffold has remained a contentious issue. Irrespective of the ongoing debate, all identified scaffold constituents must fulfil one or multiple roles of their own.

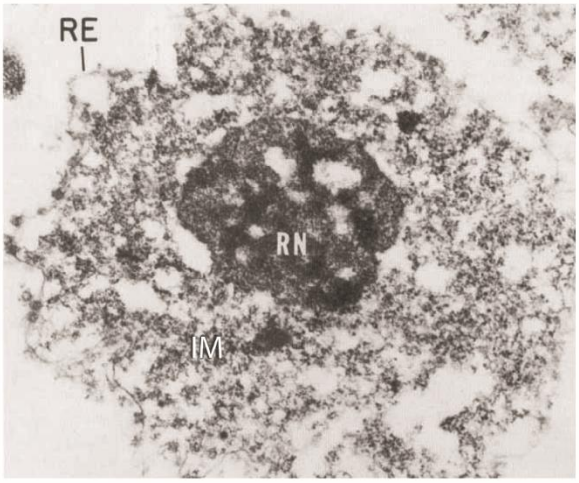


Figure 1.6 – The nuclear matrix in electron microscopy.

Original whole mount electron microscopic image by Berezney (1991) showing an isolated rat liver nuclear matrix with internal matrix (IM), residual nuclear envelope or nuclear lamina (RE), and residual nucleoli (RN) (Berezney 1991).

Figure 1.6 shows that isolated nuclear matrices maintain the principal features of intact nuclei despite the removal of 75-90% of total nuclear protein (Berezney 1991). The major components of these matrices are proteins and tightly bound RNAs. Nuclear matrices are made of a large variety of polypeptides with an enrichment of high molecular weight, non-histone species (Berezney 1991). Over 200 protein constituents have been identified with two-dimensional gel electrophoresis, many of these proteins are nucleic acid-binding (Fey, Pflug et al. 1988, Stuurman, Meijne et al. 1990). Three of the major proteins are the lamins A, B and C that constitute the remainders of the nuclear envelope (Fey, Pflug et al. 1988).

Along with those, nuclear matrins and heterogeneous ribonucleoproteins (hnRNPs) were found. The latter were identified in studies utilizing immunopurification experiments with monoclonal antibodies followed by two-dimensional gel electrophoresis (Dreyfuss, Choi et al. 1984).

1.3.1 The nuclear matrix/scaffold is a controversial structure

As mentioned in the introductory section 1.3, there is a lot of controversy around the nuclear scaffold/matrix. Formally, a protein that resists nuclear extraction is part of the nuclear matrix. However, there is a multitude of different extraction methods that are essentially incomparable (Wilson and Coverley 2013). The use of non-physiological conditions, especially high-salt buffers, for matrix extractions has been heavily criticized because these conditions may induce artificial protein aggregation (Pederson 1998).

Interestingly, a “true” nuclear scaffold can only be seen in electron microscopic images. Immunofluorescence images of scaffold proteins, however, show only a spotty distribution of the target protein(s) (Wilson and Coverley 2013).

Moreover, it seems impossible to visualize the matrix in live cells – a finding that strongly argues against the existence of such a rigid structure in the living organism (Hancock 2000). On the contrary, many matrix proteins have been shown to be rather mobile in the living cell (Hancock 2000). The only true filamentous structures that have been observed within the nucleus are composed of actin proteins and lamins. They form a fine fibrous network within the nucleus, which is not equivalent to the large structures seen in nuclear matrix preparations (Hancock 2000).

In addition to that there seems to be a discrepancy between the different model systems used to visualize the nuclear matrix, as recent research suggests that the nuclear matrix architecture changes in the course of cellular differentiation or during transformation (Wilson and Coverley 2013). Therefore, depending on the cellular system used, experimental results may show a lot of variation and the generalized term nuclear matrix/scaffold has to be used with caution (Munkley, Copeland et al. 2011).

1.3.2 Heterogeneous ribonucleoproteins – Eclectic RNA binding proteins

HnRNPs are typically associated with heterogeneous nuclear RNA (hnRNA). Together, proteins and hnRNA, form large hnRNP complexes that are thought to be involved in RNA processing (Dreyfuss, Choi et al. 1984). Initial studies identified about 20 major hnRNPs that were named alphabetically from A to U (c.f. Figure 1.7) (Pinol-Roma, Choi et al. 1988). But there are many more putative *hnrnp* factors that remain to be characterized (Carpenter, MacKay et al. 2006). Every hnRNP contains at least one RNA binding motif, viz. a RNA recognition motif (RRM), a hnRNP K homology domain (KH) or an arginine/glycine-rich box (RGG box) (Dreyfuss, Matunis et al. 1993, Krecic and Swanson 1999). Oftentimes these motifs display a preference for specific nucleic acid sequences (Beyer, Miller et al. 1980, Steitz and Kamen 1981, Bennett, Pinol-Roma et al. 1992, Matunis, Matunis et al. 1993). In addition to RNA binding domains, some hnRNPs also contain amino acid sequences that mediate protein-protein or DNA-protein interactions (Dreyfuss, Matunis et al. 1993, Cartegni, Maconi et al. 1996, Krecic and Swanson 1999). Correlating with the variety of hnRNPs identified are the cellular functions ascribed to them. Amongst others, hnRNPs play a role in DNA replication, transcription, the processing of primary transcripts, and in nuclear export, subcellular localization translation and stabilization of mRNA (Krecic and Swanson 1999, Dreyfuss, Kim et al. 2002, Glisovic, Bachorik et al. 2008).

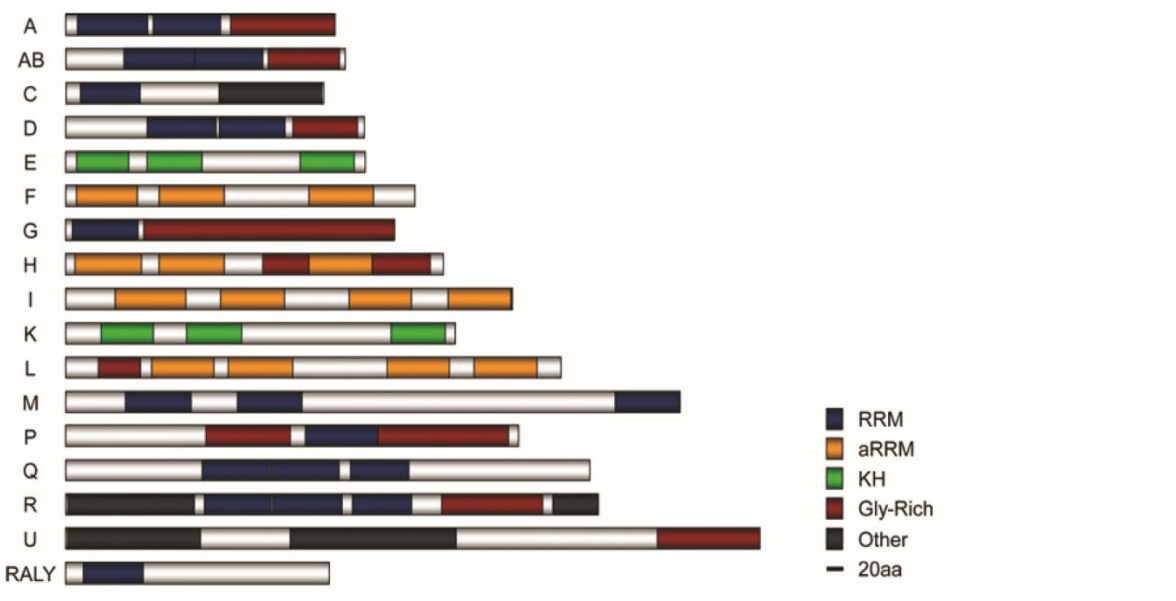


Figure 1.7 – Schematic domain structures of major hnNRPs.

Graphic retrieved from Han *et al.* (2010) showing the domain structure of the classical hnRNPs. Note, that the first polypeptide (A) represents the proteins hnRNP A0, A1, A2/B1 and A1 as they have similar domain structures. Individual domain names are abbreviated as follows: RNA recognition motif (RRM), aRRM (atypical RRM), hnRNP K homology domain (KH), and Gly-rich (glycine-rich) (reviwed in Han, Tang et al. 2010).

Though hnRNPs fulfill many different functions, they also have some common features. Not only were they initially described as components of the same complex, but they also have shared structural characteristics. With the exception of SAF-A/hnRNP U, they all carry a classical RNA binding domain (RBD), sometimes even in tandem (Figure 1.7) (reviewed in Han, Tang et al. 2010). This leads to one of the most evident properties of hnRNPs that is their RNA binding behavior. Initial *in vitro* binding assays showed that hnRNPs bind to a variety of RNA and single-stranded DNA (ssDNA) substrates in a salt- and heparin-resistant manner (Dreyfuss, Matunis et al. 1993). Later assays, under more stringent washing conditions, and Systematic Evolution of Ligands by Exponential Enrichment (SELEX) experiments showed that each of the hnRNPs had binding preferences for specific subsets of sequences (Swanson and Dreyfuss 1988, Tuerk and Gold 1990). Nowadays, hnRNP

binding specificities are often determined using the rigorous, but also technically challenging technique *in vivo* crosslinking and immunoprecipitation (CLIP) that is discussed in more detail in Chapter 3 of this thesis (Xiao, Tang et al. 2012, Rossbach, Hung et al. 2014).

The focus of this thesis work is on one particular hnRNP that has been put into the context of XCI. SAF-A/hnRNP U is an RGG box-containing hnRNP and will be described in more detail in the following Section (Section 1.3.2).

1.3.3 SAF-A/hnRNP U – A DNA and RNA binding nuclear matrix protein

In 1992, HeLa cell SAF-A/hnRNP U was discovered by two independent research groups (Kiledjian and Dreyfuss 1992, Romig, Fackelmayer et al. 1992). With a molecular weight of 120 kDa it was the largest hnRNP identified in these initial studies. Interestingly, SAF-A/hnRNP U's observed molecular weight deviates from the calculated theoretical value of about 89 kDa (Kiledjian and Dreyfuss 1992). Human and murine SAF-A/hnRNP U have similar sequence motifs and domain structures (for more detail and illustrations see results Chapters 2 and 3): They both accommodate a DNA binding SAF box at the N-terminus of the protein and a C-terminal, RNA binding RGG box. In between there is a spIA/Ryanodine receptor domain (SPRY domain), which may be involved in protein multimerization (Fackelmayer, Dahm et al. 1994, Spraggon, Dudnakova et al. 2007). Currently, there is no full crystal structure of SAF-A/hnRNP U available, which is probably due to the fact that the protein contains many inherently unstructured regions (personal communication with Dr. James Smith, University of Cambridge). Therefore, it is difficult to make *in silico* predictions on binding pocket structures, and to computationally design site-directed mutagenesis strategies.

The DNA binding property of SAF-A/hnRNP U was one of the earliest observed features (Romig, Fackelmayer et al. 1992). From then on it was believed that the protein preferentially locates to Scaffold/Matrix Attachment Regions (S/MARs) that are DNA elements in the eukaryotic genome known to mediate the interaction between chromatin and

the nuclear scaffold (Liebich, Bode et al. 2002). On average they are 600 to 3,000 bp long and locate to non-random regions within the genome, viz. close to enhancers, 5' introns and telomeres and gene breakpoint clusters (Cockerill and Garrard 1986, Gasser and Laemmli 1986, Bode, Schlake et al. 1995). S/MARs are believed to structure the genome into functional domains by facilitating chromatin loop formation. Interestingly, they do not have an apparent consensus sequence, and their only common feature is their relative AT-richness (Liebich, Bode et al. 2002). Therefore, it is thought that S/MARs are being recognized by their structural properties and/or the clustering of short sequence motifs that are not present in non-S/MAR elements (Kipp, Gohring et al. 2000). Hence, S/MAR function cannot be determined by primary sequence patterns, and for the very same reason only a minor proportion of S/MARs has been annotated to date (Liebich, Bode et al. 2002). This is why attempts to computationally predict S/MARs have been made (Girod, Nguyen et al. 2007). According to these predictions there is an enrichment of S/MARs on the X chromosome compared to autosomes. Even more striking is the discovery that LINEs and S/MARs often overlap, and in humans LINE sequences are twofold overrepresented in S/MAR regions relative to the overall LINE abundance in the genome (Jordan, Rogozin et al. 2003, Purbowasito, Suda et al. 2004).

The second observation made in early studies on SAF-A/hnRNP U, was the protein's ability to bind RNA (Kiledjian and Dreyfuss 1992). More specifically, it was found that SAF-A/hnRNP U binds to pre-mRNA *in vivo* and RNA and to single-stranded DNA *in vitro*. A recent CLIP study (2012) in HeLa cells uncovered putative RNA recognition motifs for the human protein (Xiao, Tang et al. 2012). Furthermore, associations with the regulatory, non-coding RNAs HTR, 7SK, NEAT 1/2 and HOTAIR were proposed. However, the RNA-association partners for SAF-A/hnRNP U in mouse cells are unknown.

Thirdly, SAF-A/hnRNP U was shown to associate with a variety of proteins, indicating that it fulfills more than one role in the cell nucleus (discussed in detail in Section 2.3 of Chapter 2).

On the organismal level SAF-A/hnRNP U was found to play a role in brain development. In patients with a deletion of the *HNRNPU* gene mental and motor retardation, delayed speech development, hypotonia and (febrile) seizures were observed (Caliebe, Kroes et al. 2010). Also in the mouse model the lack of an intact *safa/hnrnpu* gene severely affected development. In this context it was demonstrated that a mutation induced by the insertion of the U3Neo gene trap retrovirus into an intron of the *hnrnpu* gene caused developmental abnormalities after 6.5 *dpc*, and embryos were resorbed around day 10.5 (Roshon and Ruley 2005). Furthermore, expression of the inserted neomycin-resistance gene involved an alternative splicing event resulting in a hypomorphic *hnrnpu* mutation. Homozygous, hypomorphic mutant cell lines were then isolated from pre-implantation blastocysts. These expressed SAF-A/hnRNP U at two to five times lower levels than wild type cells. The findings imply that full levels of SAF-A/hnRNP U are required for faithful development, thus, indicating that the protein is indispensable for mammalian nuclear function.

1.3.4 SAF-A/hnRNP U in X chromosome inactivation

Recent studies linked SAF-A/hnRNP U to XCI by showing that in female human cell lines GFP-tagged SAF-A/hnRNP U is particularly enriched at the Xi (Helbig and Fackelmayer 2003, Pullirsch, Hartel et al. 2010). Moreover, deletion analyses revealed that removal of either the SAF box or the RGG domain abrogate XIST/Xist RNA homing to the Xi (Hasegawa, Brockdorff et al. 2010). Another experimental approach, involving the exogenous expression of Xist, led to the *de novo* accumulation of SAF-A/hnRNP U (Pullirsch, Hartel et al. 2010). Interestingly, SAF-A/hnRNP U accumulation on the Xi was only detected at later stages of XCI and not during initial Xist RNA accumulation. This discrepancy can only be explained by the assumption that SAF-A/hnRNP U accumulation at the Xi is a step-wise process (Hasegawa and Nakagawa 2011). This means that at initial stages of XCI SAF-A/hnRNP U interacts with only small amounts of Xist RNA that are just enough to trigger XCI. Since SAF-A/hnRNP U is a critical component of the nuclear scaffold

that is broadly distributed throughout the nucleus, an enrichment of the protein at the Xi may not be visible at this stage. Later on, greater Xist-SAF-A/hnRNP U multimer-complexes may be recruited to the Xi at a time, leading to detectable SAF-A/hnRNP U accumulation (Hasegawa and Nakagawa 2011).

Hasegawa *et al.* (2010) made a first attempt to reveal a functional relationship between SAF-A/hnRNP U and Xist RNA (Hasegawa, Brockdorff et al. 2010). They took a small interfering RNA (siRNA) screening approach showing that in cells treated with siRNAs against SAF-A/hnRNP U Xist failed to coat the Xi. Instead, Xist molecules were found dispersed within the nucleus. Furthermore, SAF-A/hnRNP U depletion in differentiating ESCs interfered with XCI in these cells (Hasegawa, Brockdorff et al. 2010). To complement their *in vivo* data, Hasegawa *et al.* conducted UV crosslinking RNA immunoprecipitation (RNA IP) experiments in cells bearing full-length and individual domain deletion constructs of SAF-A/hnRNP U. Based on their results they put forward the idea that the RGG box of SAF-A/hnRNP U binds directly to Xist RNA such that the lncRNA is brought to the Xi together with other silencing factors like PRC2 components (Figure 1.8).

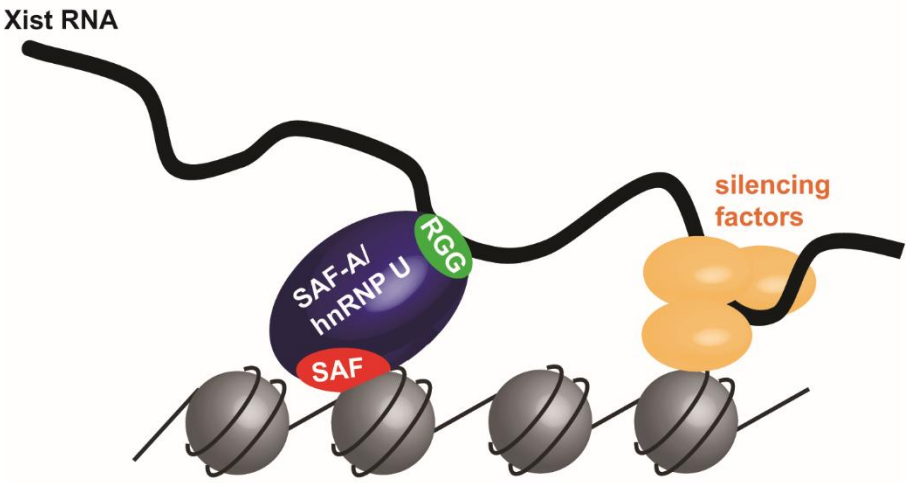


Figure 1.8 – Model for SAF-A/hnRNP U action at the Xi proposed by Hasegawa *et al.* (2010).

SAF-A/hnRNP U binds to Xist RNA via its RNA binding RGG domain. Then the protein-RNA complex is recruited to the Xi. On its way Xist also associates with silencing factors (e.g. PRC2 proteins) that entail X chromosome compaction and transcriptional silencing.

Notably, not all observations made by Hasegawa *et al.* were coherent. For instance, they found that SAF-A/hnRNP U knock down effects varied greatly between cell lines. Whereas they did not observe a great extent of Xist RNA mislocalization in mouse embryonic fibroblasts, the mouse neuroblastoma cell line Neuro2a showed clear results. Moreover, no enrichment of H3K27me3 at the Xi of siRNA-treated Neuro2a cells was observed (Hasegawa, Brockdorff et al. 2010). Consequently, Hasegawa's results may only show a cell type-specific mechanism for Xist RNA recruitment and not be generally applicable to other cellular systems.

In summary, more research needs to be done to determine the exact molecular role for SAF-A/hnRNP U in XCI. Specifically its relationship to Xist RNA and Xist-mediated silencing need to be investigated further. The present thesis ties into the body of literature at this point.

1.4 Aims of this thesis

Previously, Xist RNA was shown to remain associated with the nuclear scaffold, even after undergoing the stringent matrix extraction procedure (Clemson, McNeil et al. 1996). Because of this finding SAF-A/hnRNP U has moved into the focus of X inactivation researchers.

The present study was designed to re-investigate SAF-A/hnRNP U's RNA binding behavior in mouse ESCs carrying an inducible *Xist* gene and determine its exact RNA binding consensus sequence using CLIP-Seq. Thus, confirming or disproving the previously suggested, direct SAF-A/hnRNP U – Xist RNA binding (Hasegawa, Brockdorff et al. 2010). This work was complemented with investigations of the SAF-A/hnRNP U interactome in

these cells, possibly highlighting a direct connection between SAF-A/hnRNP U and Xi silencing. In conjunction with the interactome studies, an attempt was made to identify the reason(s) for the stable and increased local SAF-A/hnRNP U concentration at the Xi (Fackelmayer 2005). This research was extended by looking at SAF-A/hnRNP U during the cell cycle of somatic mouse cell lines, and trying to correlate the protein's dynamics with Xist RNA turnover.

The experimental results of these endeavors are presented in the following Chapters (2 to 6) and will give an impression of what role(s) in XCI may be fulfilled by SAF-A/hnRNP U.

2. SAF-A/hnRNP U is a pan-nuclear scaffolding protein

2.1 Introduction

The nuclear scaffold/nuclear matrix is a subcellular structure that remains after treating nuclei with high-salt buffers and nucleases. It is a fibrous network of proteins with only residual amounts of DNA, RNA, carbohydrates and phospholipids (Berezney and Coffey 1977). Even though much research has been done on the individual constituents of the nuclear scaffold, the function of the structure as such has remained elusive (Pederson 2000). The present work, however, does not aim at elucidating the cellular role of the nuclear matrix. Instead, it discusses the molecular function of a specific nuclear scaffold component and its role in the process of mammalian XCI. The protein of interest is the Scaffold Attachment Factor-A (SAF-A) that for historical reasons was also given the name heterogeneous ribonucleoprotein U (hnRNP U) (Kiledjian and Dreyfuss 1992, Romig, Fackelmayer et al. 1992).

SAF-A/hnRNP U is a scaffolding protein with an apparent molecular weight of 120 kDa (Dreyfuss, Matunis et al. 1993). Previous studies revealed that only one half of the cellular pool of nuclear SAF-A/hnRNP U is bound to the nuclear scaffold (Fackelmayer and Richter 1994, Mattern, Humbel et al. 1996, Mattern, van der Kraan et al. 1999). The other half is divided equally between a soluble hnRNP-bound and a DNase-extractable fraction. In fact, at least 20 % of nuclear SAF-A/hnRNP U was found to bind to chromosomal DNA (Romig, Fackelmayer et al. 1992, Fackelmayer, Dahm et al. 1994, Gohring and Fackelmayer 1997). Moreover, it was shown by electron microscopy that the isolated protein binds to both single- and double-stranded DNA and forms higher-order molecular structures. To facilitate DNA binding, SAF-A/hnRNP U has a dedicated DNA binding domain (SAF box). The SAF box is present in many other proteins in a variety of species and it is structurally related to a homeodomain that is a type of DNA or RNA binding domain usually found in transcription factors (Gehring 1992, Kipp, Gohring et al. 2000). Besides the SAF box, an arginine- and

glycine-rich RNA binding domain (RGG domain) and a spIA/Ryanodine receptor domain (SPRY domain) have been identified. The latter may be involved in protein-protein interactions and SAF-A/hnRNP U multimerization (Fackelmayer, Dahm et al. 1994, Ponting, Schultz et al. 1997, Spraggon, Dudnakova et al. 2007). However, up to now no clear evidence associating the SPRY domain with this function has been published.

Apart from being a constituent of the nuclear scaffold, more recently SAF-A/hnRNP U was shown to associate with the inactive X chromosome (Xi) in female cells. In 2003, Helbig and Fackelmayer demonstrated that enhanced green fluorescent protein (eGFP)-tagged SAF-A/hnRNP U is locally enriched at the Xi in interphase nuclei of human cells (Helbig and Fackelmayer 2003). Later this finding was substantiated by Pullirsch *et al.* (2010) who replicated the results and identified a human autoimmune serum that highlighted SAF-A/hnRNP U enrichment at the Xi (Pullirsch, Hartel et al. 2010). At the same time Hasegawa *et al.* (2010) performed SAF-A/hnRNP U knockdown experiments revealing that SAF-A/hnRNP U depletion perturbs X inactive-specific transcript (Xist) RNA localization. Furthermore, using RNA-immunoprecipitation (RNA-IP) they provided evidence that SAF-A/hnRNP U specifically binds to Xist RNA (Hasegawa, Brockdorff et al. 2010).

Although using fluorescently tagged SAF-A/hnRNP U shows enrichment at the Xi, none of the antibodies raised against endogenous SAF-A/hnRNP U detected the Xi-bound form under normal indirect immunofluorescence (IF) conditions. A single study observed human SAF-A/hnRNP U at the Xi by IF imaging, but only after a classical *in situ* nuclear matrix preparation (Helbig and Fackelmayer 2003). Notably, these studies were done on human cells only. Because oftentimes human and mouse cells behave differently and the focus of this study will be on the mouse protein, raising an antibody against murine SAF-A/hnRNP U was the first step in this study.

2.2 Generation of an antibody specifically recognizing SAF-A/hnRNP U

To have a reliable reagent for IF detection and biochemical applications, a polyclonal antibody against murine SAF-A/hnRNP U was raised. After examining the Jameson-Wolf antigenic index plot for SAF-A/hnRNP U, the 260 most N-and C-terminal amino acid residues, amino acids 1 to 260 and 540 to 800 respectively, were identified as suitable antigens for immunization (Jameson and Wolf 1988). Both regions were cloned, recombinantly expressed in *E. coli* and affinity purified on Ni-NTA resin. Whereas the N-terminal polypeptide was soluble (Figure 2.1B), attempts to purify soluble C-terminal SAF-A/hnRNP U using various epitope tags (His-tag, GST-tag and MBP-tag) failed. Figure 2.1A shows a schematic of full length mouse SAF-A/hnRNP U and the selected N-terminal antigen used to immunize the rabbit.

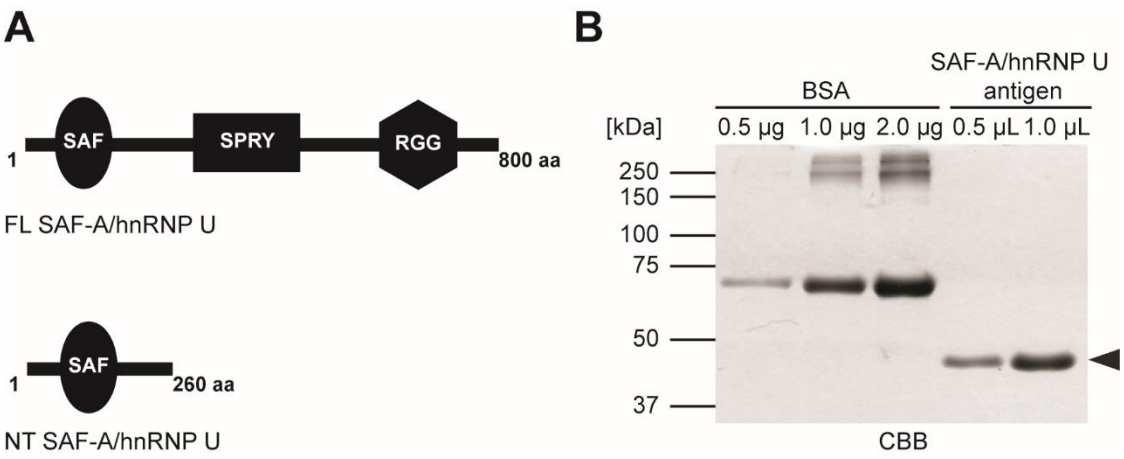


Figure 2.1 – Antigen design and purification.

(A) Schematic representation of the full length (FL) protein (amino acids 1 to 800) and the N-terminal (NT) peptide (amino acids 1 to 260) used for immunization of the rabbit. (B) Coomassie Brilliant Blue G-250 (CBB) stained acrylamide gel with purified SAF-A/hnRNP U antigen marked by the arrowhead (▼) and a bovine serum albumin (BSA) standard (~66.5 kDa).

One pre-immunization and three consecutive post-immunization bleeds were taken from a single rabbit. Upon receipt, the crude antisera were tested for reactivity against

recombinantly expressed and endogenous SAF-A/hnRNP U. With the exception of the pre-immune serum, all bleeds tested showed reactivity towards recombinant and endogenous SAF-A/hnRNP U protein of human and mouse origin in Western blot analysis (Figure 2.2).

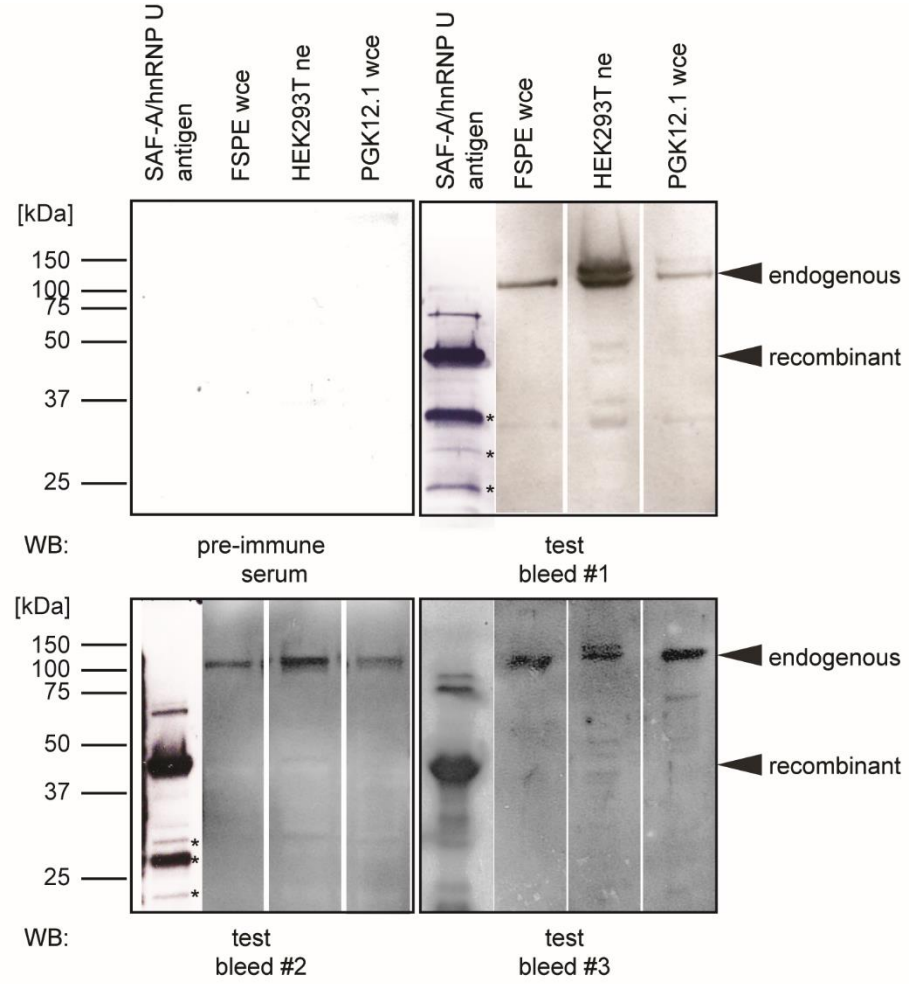


Figure 2.2 – Anti-SAF-A/hnRNP U recognizes the recombinant fragment and endogenous protein.

Western blot (WB) analysis using pre-immune serum and three consecutive bleeds collected after immunization with the SAF-A/hnRNP U antigen. Antibody reactivity was tested on the original antigen purified from *E.coli* (~40 kDa), and on the endogenous protein (~120 kDa) using whole cell extract (wce) from female *Mus spretus* ear (FSPE) tissue-derived cells, nuclear extract (ne) from human embryonic kidney cells (HEK293T), and wce of mouse embryonic stem cells (PGK12.1). WB bands showing the endogenous and the recombinant

proteins are marked by arrowheads (▼). Protein degradation products in the SAF-A/hnRNP U antigen lanes are indicated by asterisks (*).

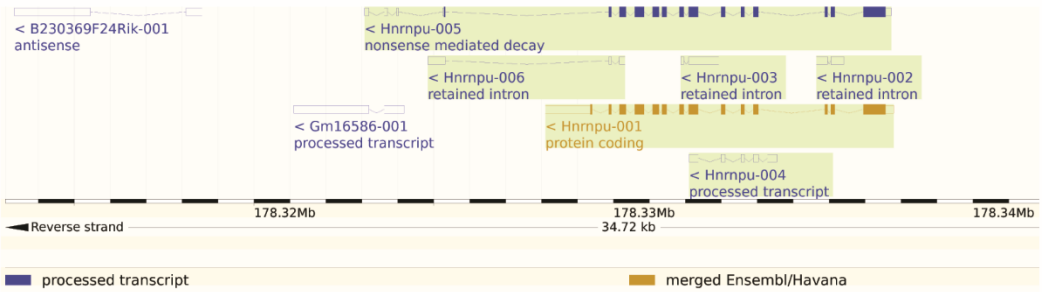
Lanes with recombinant antigen yielded additional bands, predominantly at a lower molecular weight than the full length antigen. These likely correspond to protein degradation products (Figure 2.2, marked by asterisks). In contrast, analysis of whole cell and nuclear extract revealed very weak cross-reactive bands, indicating that the antiserum has high specificity for SAF-A/hnRNP U.

In some instances a doublet was observed for SAF-A/hnRNP U (Figure 2.2 and Figure 2.4B). The most plausible explanation for this would be targeted proteolysis of SAF-A/hnRNP U, evidence for which has been published previously (Gohring, Schwab et al. 1997). In their study Göhring *et al.* (1997) observed two SAF-A/hnRNP U species after caspase-dependent protein cleavage. Whereas the full length protein was observed at a molecular weight of about 120 kDa the cleavage product was found at 97 kDa. Göhring *et al.* also put forward the idea that non-cleaved protein associates with DNA. Once the protein has been cleaved into its shorter form, it is released from the DNA and the nuclear scaffold. The cleavage event was shown to be concomitant with apoptosis (Gohring, Schwab et al. 1997).

An alternative explanation is the existence of two distinct isoforms, suggested by Fackemayer and Richter in 1994 (Fackelmayer and Richter 1994). However, there are no further studies following up on this report in the published literature. In contrast, the Ensembl database lists several splice forms for both the human and the mouse protein (Figure 2.3). In human these give rise to three putative protein products. For mouse two mRNA isoforms have been identified through the Human and Vertebrate Analysis and Annotation (HAVANA) project that aims at manually annotating human, mouse and zebrafish genomes. One is the protein investigated in this study (Hnrnpu-001). The other (Hnrnpu-005) may not give rise to an actual protein product.

mouse SAF-A/hnRNP U

Name	Transcript ID	Length [bp]	Protein ID	Length [aa]	Biotype
Hnrnpu-001	ENSMUST00000037748	3878	ENSMUSP00000047571	800	Protein coding
Hnrnpu-005	ENSMUST00000161769	2844	ENSMUSP00000124147	793	Nonsense mediated decay
Hnrnpu-004	ENSMUST00000146362	869	No protein product	-	Processed transcript
Hnrnpu-003	ENSMUST00000150825	923	No protein product	-	Retained intron
Hnrnpu-006	ENSMUST00000162916	710	No protein product	-	Retained intron
Hnrnpu-002	ENSMUST00000134291	673	No protein product	-	Retained intron



human SAF-A/hnRNP U

Name	Transcript ID	Length [bp]	Protein ID	Length [aa]	Biotype
HNRNPU-001	ENST00000444376	5940	ENSP00000393151	806	Protein coding
HNRNPU-002	ENST00000283179	3704	ENSP00000283179	825	Protein coding
HNRNPU-003	ENST00000440865	1067	ENSP00000410728	242	Protein coding
HNRNPU-004	ENST00000366525	2535	No protein product	-	Retained intron
HNRNPU-008	ENST00000476241	1687	No protein product	-	Retained intron
HNRNPU-007	ENST00000483966	1438	No protein product	-	Retained intron
HNRNPU-005	ENST00000465881	741	No protein product	-	Retained intron
HNRNPU-006	ENST00000468690	561	No protein product	-	Retained intron

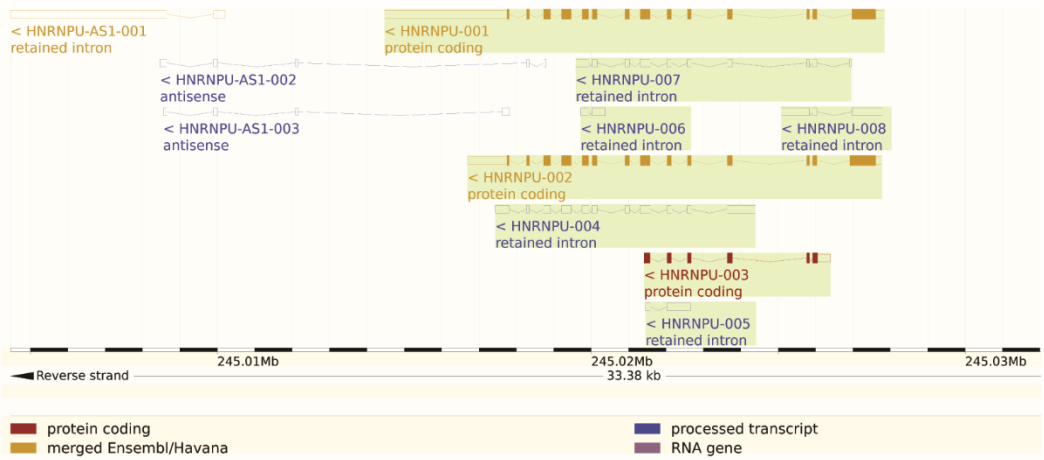


Figure 2.3 – Ensembl database entries for mouse and human SAF-A/hnRNP U.

Tables and screenshot images representing the Ensembl database output for mouse (top) and human (bottom) SAF-A/hnRNP U. The tables provide information on the transcript name, identifier (ID), transcript length in base pairs (bp), the length of the protein product in amino acids (aa) and the biotype. On the database there are six and eight spliceoforms for mouse

and human SAF-A/hnRNP U, respectively. In mouse, only two of them may be protein-coding. One out of the two (Hnrnpu-005) is believed to be subject to nonsense mediated decay at mRNA level, the other is the protein investigated in this study (highlighted in yellow). The human splice forms give rise to three protein products.

An additional affinity purification step was included to improve antibody specificity, and to produce a consistent reagent for future applications. All bleeds were immunoaffinity purified using the immunization antigen covalently coupled to an activated immunoaffinity agarose support matrix (AffiGel). The antibody was eluted at low pH and immediately neutralized to avoid irreversible unfolding of the protein (for more details see Materials and Methods Section). Coomassie staining of a sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) gel revealed two discernible bands corresponding to the IgG heavy and light chain (Figure 2.4A). Individual fractions were pooled post purification and tested in Western blotting (Figure 2.4B).

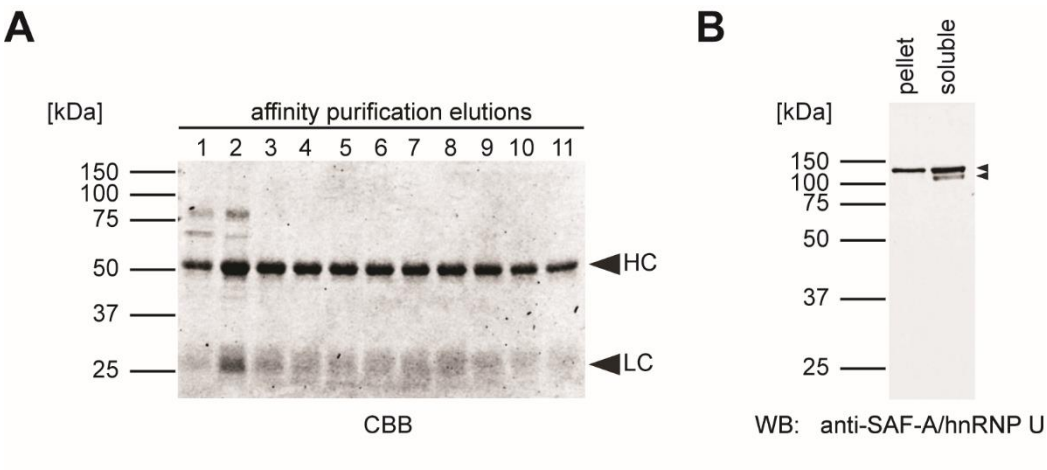


Figure 2.4 – Affinity purification further reduces anti-SAF-A/hnRNP U background signal.

(A) CBB-stained sodium dodecyl sulfate - polyacrylamide gel electrophoresis (SDS-PAGE) gel with anti-SAF-A/hnRNP U fractions affinity purified from crude serum (elution 1 through 11). IgG heavy (HC) and light chain (LC) are indicated by arrowheads (▼). (B) Western blot on pellet and soluble fraction of FSPE nuclear extract preparation using affinity purified anti-SAF-A/hnRNP U. The arrowheads (▼) pinpoint the two SAF-A/hnRNP U bands.

Importantly, the antibody was also tested in IP applications. Figure 2.5 shows that anti-SAF-A/hnRNP U specifically precipitates endogenous SAF-A/hnRNP U protein from mouse embryonic stem cell (ESC) nuclear extract. In contrast, residual SAF-A/hnRNP U was seen in the IgG control IP (Figure 2.5B). It is very likely that the SAF-A/hnRNP U background band observed in the IgG control lane (Ctrl.) of anti-SAF-A/hnRNP U Western blot is due to the fact that SAF-A/hnRNP U is a very abundant nuclear protein. This abundance makes it more likely that non-specific interactions are recovered with the immunoaffinity matrix (Markham, Bai et al. 2007).

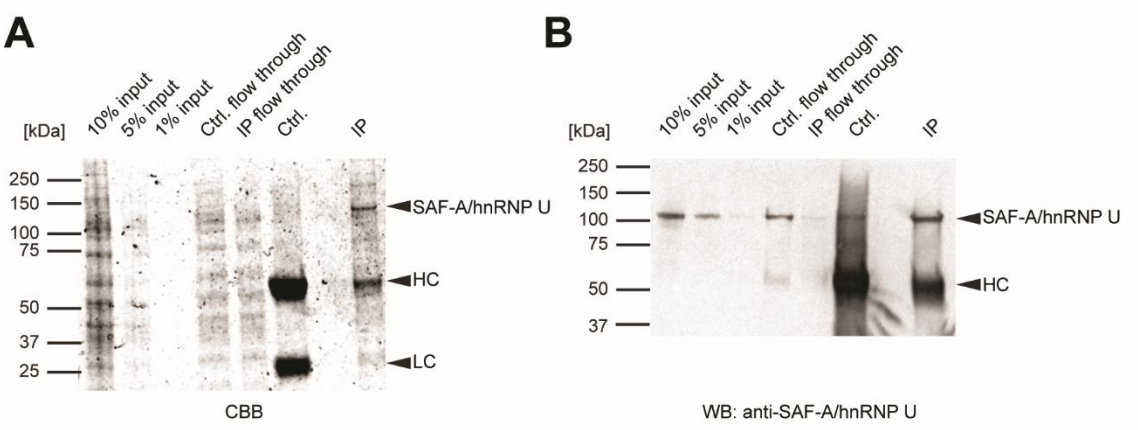
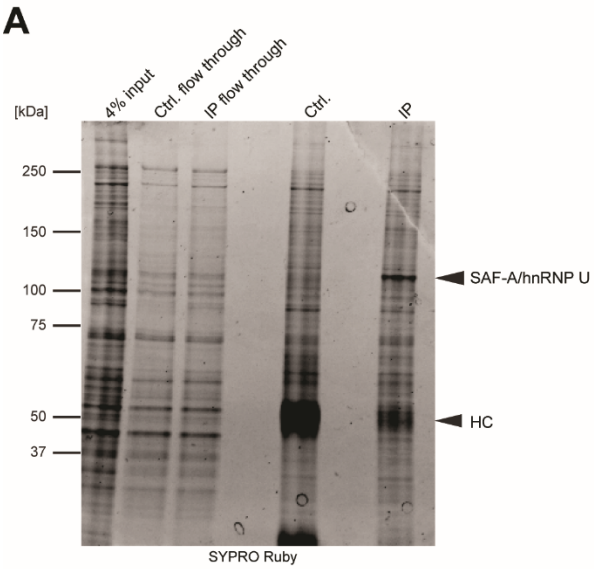


Figure 2.5 – SAF-A/hnRNP U immunoprecipitation.

(A) CBB stained SDS-PAGE gel of IgG control (Ctrl.) and anti-SAF-A/hnRNP U (IP) IP samples on 100 µg P6D7 mouse embryonic stem cell (ESC) nuclear extract. SAF-A/hnRNP U and the antibody HC and LC are indicated by arrowheads (▼). (B) Western blot analysis with anti-SAF-A/hnRNP U on IgG control (Ctrl.) and anti-SAF-A/hnRNP U (IP) IPs on 100 µg P6D7 ESC nuclear extract. Again SAF-A/hnRNP U and the antibody HC are indicated by arrowheads (▼).

In order to validate the IP experiment and to exclude that the pulled down protein is a contaminant with the same apparent molecular weight (~120 kDa) as SAF-A/hnRNP U, an IP band was cut off the gel and submitted for mass spectrometry (MS) (Figure 2.6). Subsequent Mascot analysis confirmed that SAF-A/hnRNP U was enriched in the IP sample (Figure 2.6C). On the contrary, no SAF-A/hnRNP U was found in the IgG control sample.

Altogether, these results suggest that the SAF-A/hnRNP U antibody prepared for this study is a reliable tool for biochemical applications.



B

1	MSSSPVNVKK	**LKVSELKEEL**	**KK**RRL**SDKGL**	**KADLMDRLQA**	**ALDNEAGGRP**
51	**AMEPGNGSLD**	**LGGDAAGR**SG	AGLEQEAAAG	AEDDEEEEGI	AALDGDQMEL
101	GEENGAAGAA	DAGAMEEEEA	ASEDENGDDQ	GFQEGEDELG	DEEEGAGDEN
151	GHGEQQSQPP	AAAAQQQPSQ	QRGAGKEAAG	K**SSGPTSLFA**	**VTVAPPGAR**Q
201	GQQQAGGDGK	TEQKGGDKKR	GVKRPR**EDHG**	**RGYFEYIEEN**	**KYSR**AKSPQP
251	PVEEEDEHFD	DTVVCLDTYN	CDLHFKISRD	RLSASSLTME	SFAFLWAGGR
301	ASYGVSK**GKV**	**CFEMKVTEKI**	**PVRHLYTKDI**	**DIHEVR**IGWS	LTTSGMLLGE
351	EEFSYGYSLK	GIK**TCNCETE**	**DYGEK**FDEND	VITCFANFET	DEVELSYAK**N**
401	**GQDLGVAFKI**	**SK**EVLADRPL	FPHVLCHNCA	VEFNFGQK**EK**	**PYFPIPEDCT**
451	**FIQNVPLEDR**	VRGPK**GPEEK**	**KDCEVVMMIG**	**LPGAGK**TTWV	TK**HAAENPGK**
501	**YNILGTNTIM**	**DKMMVAGFKK**	**QMADTGKLNT**	**LLQRAPQCLG**	**KFIEIAARKK**
551	**RNFILDQTNV**	**SAAAQRRKMC**	**LFAGFQRKAV**	**VVCPKDEDYK**	**QRTQKKAEVE**
601	**GKDLPEHAVL**	**KMKGNFTLPE**	**VAECFDEITY**	**VELQKEEAQK**	**LLEQYKEESK**
651	**KALPPEKKQN**	**TGSK**KSNKNK	SGKNQFNRGG	GHRGRGGFNM	RGGNFRGGAP
701	GNR**GGYNRRG**	**NMPQRGGGGG**	**SGGIGYPYPR**	GPVFPGRGGY	SNRGNYNR**GG**
751	**MPNRGNYNQN**	**FRGR**GNNRGY	KNQSQGYNQW	QQGSVHVNVL	CEE

C

SAF-A/hnRNP U IP

Protein (gene)	Score	Cov [%]
Heterogeneous nuclear ribonucleoprotein U (Hnrnpu)	1546	50
Nucleolar RNA helicase 2 (Ddx21)	538	54
Tripartite motif-containing 28 (Trim 28)	451	59
Nucleolin (Ncl)	442	46
Splicing factor 3b, subunit 3 (Sf3b3)	441	52
Eukaryotic translation initiation factor 4 gamma 2 (Eif4g2)	382	35
Heterogeneous nuclear ribonucleoprotein U-like protein 2 (Hnrnpul2)	373	33
Putative ribosomal RNA methyltransferase NOP2 (Nop2)	367	31
LIM domain and actin-binding protein 1(Lima1)	341	37

IgG control IP

Protein (gene)	Score	Cov [%]
Keratin 10 (Krt10)	101	8
Keratin, type I cytoskeletal (Krt17)	87	6
Keratin, type II cytoskeletal 2 epidermal (Krt2)	73	3
Keratin, type II cytoskeletal 1 (Krt1)	46	4
Keratin, type II cytoskeletal 5 (Krt5)	31	1
Protein Nolc1 (Nolc1)	24	1
Serum albumin (Alb)	23	0
Golgi apparatus protein 1 (Glg1)	16	0
Keratin, type II cytoskeletal 2 oral (Krt76)	15	1

Figure 2.6 – Confirmation of SAF-A/hnRNP U specificity by mass spectrometry.

(A) SYPRO Ruby-stained 4-12% NuPAGE gel of IgG control (Ctrl.) and SAF-A/hnRNP U (IP) IP samples for mass spectrometry (MS). Arrowheads (▼) pinpoint SAF-A/hnRNP U and IgG HC. (B) Mapping of the SAF-A/hnRNP U peptides after MS analysis (50% sequence coverage). (C) List of the nine most enriched proteins in SAF-A/hnRNP U and IgG control IP. Shown are protein names, Mascot score (Score) and percent sequence coverage (Cov [%]).

2.3 SAF-A/hnRNP U associates with other RNA binding factors

Previous studies showed that SAF-A/hnRNP U interacts with a broad range of proteins. Some of these are RNA binding factors themselves; many others fulfill essential nuclear

functions. Examples include hnRNP L (Shnyreva, Schullery et al. 2000), histone acetyltransferase p300/CBP-associated factor (PCAF) (Obrdlik, Kukalev et al. 2008), p300 (Martens, Verlaan et al. 2002), Necdin (Taniura and Yoshikawa 2002), telomerase complex components (Fu and Collins 2007), RNA polymerase II (Kim and Nikodem 1999, Vizlin-Hodzic, Johansson et al. 2011), Wilm's tumor 1 (WT1) (Spraggon, Dudnakova et al. 2007), nei endonuclease VIII-like 1 (NEIL1) (Hegde, Banerjee et al. 2012), transcription activator BRG1 (Vizlin-Hodzic, Runnberg et al. 2011), Aurora A, targeting protein for Xklp2 (TPX2) (Ma, Matsunaga et al. 2011) and splicing factors (Shnyreva, Schullery et al. 2000). Notably, most interactions were identified in human cancer cell lines, specifically HeLa. Interactome studies are strongly dependent on the selected model system and the experimental conditions. The aim of this study was to look at SAF-A/hnRNP U-protein interactions in mouse ESCs and draw conclusions about its molecular function in this system. Here XY male cells that carry a doxycycline inducible Xist-transgene on chromosome 17 (3E ESCs) were examined (Tang, Huntley et al. 2010).

To identify interacting proteins, a label-free mass spectrometry (MS) approach was taken. The only difference to the IP experiments presented in the previous section, is that instead of cutting the SAF-A/hnRNP U band off the acrylamide gel, sample and IgG control were run only briefly into the gel matrix. Subsequently, the full pull-down was excised from the gel and submitted for MS. The results of three replicates were analyzed using the Central Proteomics Facility Pipeline (CPFP). Normalized spectral index quantification (SINQ), a freely available and easy to use quantitative proteomics tool, was used to identify potentially interesting SAF-A/hnRNP U interacting factors (Trudgian, Ridlova et al. 2011). Additionally, Gene Ontology (GO) searches using the MetaCore (Thomson Reuters) program were performed to classify subcellular localization, molecular functions and the biological processes associated with the SAF-A/hnRNP U binding factors (Figure 2.7).

Since SAF-A/hnRNP U localizes to the nucleus, the IPs were conducted on nuclear extracts. In the subsequent analysis the GO localization terms (Figure 2.7) clearly reflected SAF-A/hnRNP U's subcellular localization, i.e. the nucleus. For instance, SAF-A/hnRNP U is a heterogeneous ribonucleoprotein and as such it naturally occurs in ribonucleoprotein complexes (Kiledjian and Dreyfuss 1992, Romig, Fackelmayer et al. 1992). Furthermore, SAF-A/hnRNP U has been assigned a role in the regulation of alternative splicing (Xiao, Tang et al. 2012). Therefore, it is rather unsurprising that within the ten highest hits in the GO analysis the spliceosomal complex can be found under the GO term localization, and RNA splicing in the section GO processes. In the category molecular functions, RNA binding comes up first. This underscores SAF-A/hnRNP U's capability to associate with RNAs.

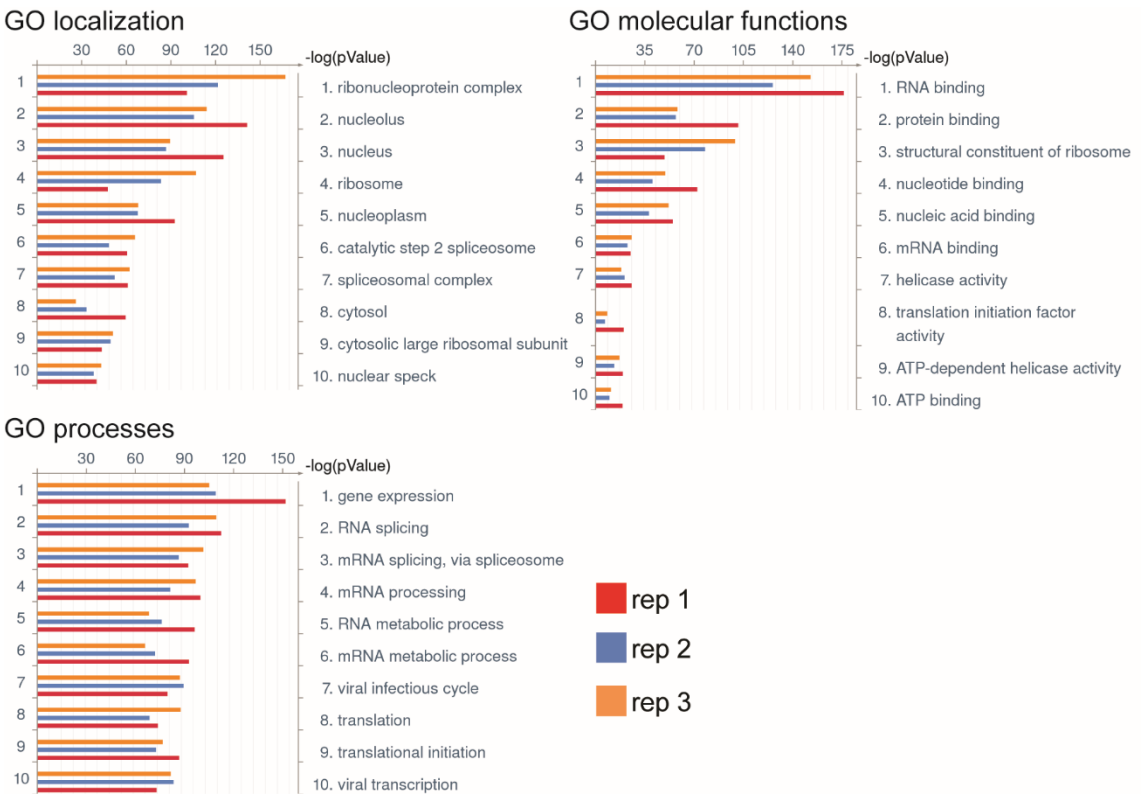


Figure 2.7 – Gene ontology analysis for the SAF-A/hnRNP U interactome.

Gene ontology (GO) analyses of the most enriched SAF-A/hnRNP U binding factors in mouse ESCs. The most enriched GO localizations, molecular functions and processes for three MS

replicates (red – rep1, blue – rep 2, orange – rep 3) based on their p-value are shown on a – log(pValue) scale.

After having assessed which GO terms apply to proteins that co-precipitate with SAF-A/hnRNP U, a closer look was taken at individual candidates. For this purpose the data were sorted into two groups: One comprised all proteins found in the sample IPs only. The other was made up of proteins which were found in both sample and control IP Figure 2.9A and B). Comparison of the three replicates showed that there was only a small overlap between the proteins found exclusively in the SAF-A/hnRNP U IP (one protein), and a more extensive overlap between the proteins found in both sample and control (359 proteins). The only common protein to all SAF-A/hnRNP U only IP samples is the Putative transposase element L1Md-A101/L1Md-A102/L1Md, short TR22. TR22 is a nucleic acid binding protein that may be involved in the retrotransposition of long interspersed elements (LINE1) within the genome (Martin and Bushman 2001).

Next, to identify the most enriched proteins found in both sample and control IPs of all three replicates were analyzed statistically. For this purpose a power law global error model (plgem) was used. This model had originally been developed for the identification of differentially expressed genes in microarray datasets and had been adapted by researchers for the determination of statistically significant protein enrichment in label-free proteomics approaches (Pavelka, Pelizzola et al. 2004). Firstly, the suitability of the model was tested: Figure 2.8 shows that plgem can be applied to the given SAF-A/hnRNP U MS dataset. Using the plgem output, a volcano plot was prepared (Figure 2.9C). This is a stringent methodology to identify significantly enriched proteins. All red data points represent proteins that were enriched in the SAF-A/hnRNP U sample, all blue data points represent proteins depleted in the sample and, thus, enriched in the IgG control IP (Figure 2.9C). The remaining black data points were below the significance threshold indicated by the dashed

lines, and therefore not included in the further analysis. As expected SAF-A/hnRNP U itself is the most enriched protein.

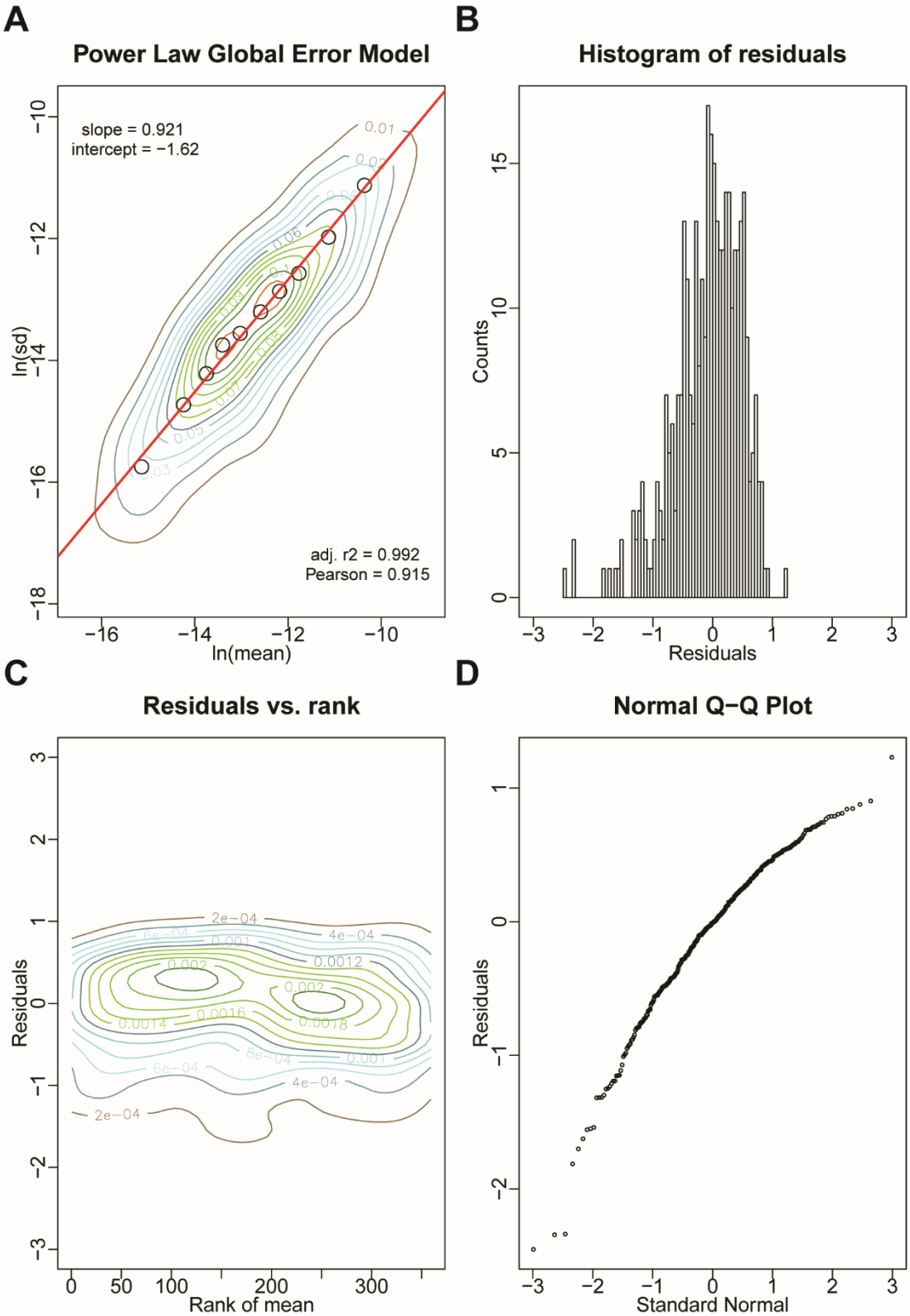


Figure 2.8 – Quality control showing suitability of power law global error model (plgem) analysis for the normalized spectral index quantification (SINQ) data.

(A) Power law global error model. This plot represents the natural logarithm of the mean (ln(mean)) versus the natural logarithm of the standard deviation (ln(sd)). The high coefficient of determination (R^2) indicates a good fit for the model, and Pearson's correlation coefficient shows that there is a positive correlation between the two variables. (B) Histogram of residuals. The histogram presents a plot of residuals versus counts and the shape approximates a Gaussian distribution with a slight left-hand skew. (C) Residuals vs. rank. This plot shows the rank of the mean plotted against the residuals. (D) Normal Q-Q plot. Quantile-quantile plot displaying standard normal values plotted against observed residuals. The plot shows that applying a normal model is a valid approach.

Also interesting is that almost all sample-enriched proteins (Table 2.1) show nucleotide binding ability, with the exception of a single contaminant that is Keratin, type 2 cytoskeletal oral (K22O). In contrast, the proteins depleted in the sample (Figure 2.9C and Table 2.1, enriched in control) are clearly very abundant background proteins, i.e. Keratin-10 (K1C10), ribosomal proteins (RPL29, RS4X and RS27) and glyceraldehyde-3-phosphate dehydrogenase (G3P).

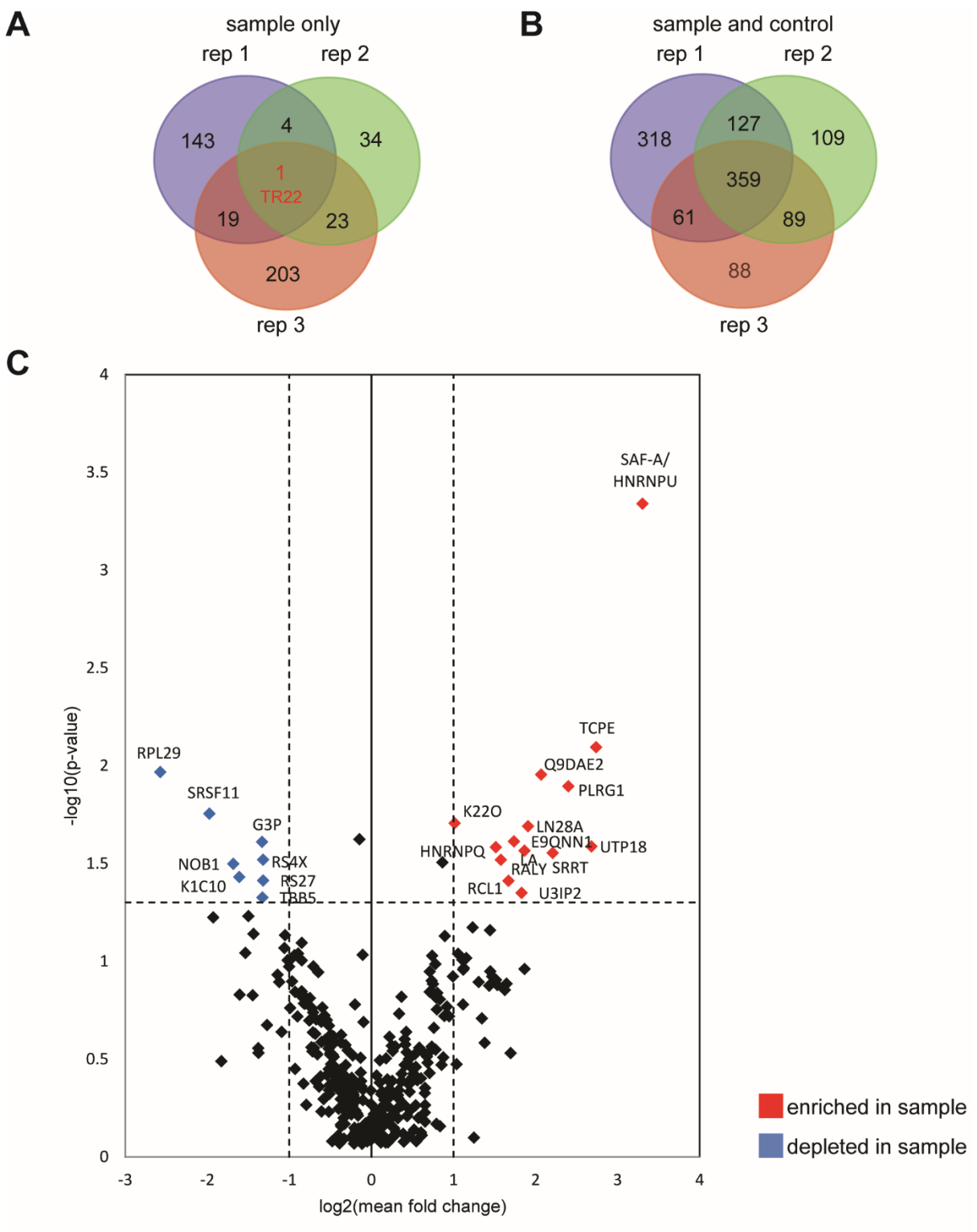


Figure 2.9 – SAF-/hnRNP U interactome analysis shows preferential association with nucleotide binding proteins.

(A) Venn diagram of proteins identified in the SAF-A/hnRNP U sample only and not the IgG control IP in the three replicates (blue – rep1, green – rep2, red – rep 3). Only one protein (TR22

- Putative transposase element L1Md-A101/L1Md-A102/L1Md) common to all samples could be identified. (B) Venn diagram of proteins identified in both SAF-A/hnRNP U sample and IgG control IP. 359 of these proteins were common to all three replicates (blue – rep1, green – rep2, red – rep 3). (C) Volcano plot for the 395 common proteins generated using the p-value output of the power law global error law (plgem) R-script. The log_2(fold change) for the replicates (n=3) is plotted against the $-log_{10}$(p-value). The cutoffs (dashed lines) were set to p-values of +/-0.05 and log_2(fold changes) of +/-1. Proteins significantly enriched in the sample versus the control IP are displayed in red. Proteins depleted in the sample a displayed in blue.

Table 2.1 – Proteins enriched and depleted in SAF-A/hnRNP U co-IPs.

Two tables showing the proteins enriched in the SAF-A/hnRNP U sample (top table) and the proteins depleted in the sample (bottom table). Full protein names and the GO terms cellular component, molecular function and biological process are shown. (nd – data not available, CRD – coding region instability determinant).

enriched in sample

Symbol	Protein name	GO cellular component	GO molecular function	GO biological process
SAF-A/HNRNPU	Scaffold attachment factor A/ Heterogeneous ribonucleoprotein U	cytoplasm, nucleoplasm, spliceosome	ATP binding, DNA binding, RNA binding	CRD-mediated mRNA stabilization, RNA splicing, mRNA processing
TCPE	T-complex protein 1 subunit epsilon	cell body, centrosome, chaperonin-containing T-complex, microtubule, zona pellucida receptor complex	ATP binding	protein folding
Q9DAE2	Rbmcl2	nucleus	nucleotide binding	nd
PLRG1	Pleiotropic regulator 1	catalytic step 2 spliceosome, nuclear speck	nd	RNA splicing and mRNA processing
LN28A	Protein lin-28 homolog A	cytoplasmic mRNA processing body, cytoplasmic stress granule, nucleolus	DNA binding, mRNA binding, miRNA binding, zinc ion binding	RNA 3' end processing, germ cell development, miRNA catabolic process, pre-miRNA processing, regulation of gene silencing by miRNA, regulation of transcription, stem cell differntiation/maintenance
UTP18	U3 small nucleolar RNA-associated protein 18 homolog	nucleolus	nd	rRNA processing
E9QNN1	ATP-dependent RNA helicase A	CRD-mediated mRNA stbility complex, centrosome, nucleolus, ribonucleoprotein complex	ATP binding, ATP-dependent helicase activity, nucleic acid-binding	CRD-mediated mRNA stabilization
K22O	Keratin, type 2 cytoskeletal oral	keratin filament	strcutural molecule activity	nd
HNRNPQ	Heterogeneous ribonucleoprotein Q	CRD-mediated mRNA stbility complex, catlytic step 2 spliceosome, histone pre-mRNA 3' end processing complex	nucleotide binding, poly(A) RNA binding	CRD-mediated mRNA stabilization, RNA splicing, cellular response to interferon gamma, mRNA processing, negative regulation of translation
LA	La ribonucleoprotein	nucleus, ribonucleoprotein complex	RNA binding, nucleotide binding	RNA processing
RALY	RNA binding protein Raly	catalytic step 2 spliceosome	RNA binding, nucleotide binding	RNA splicing, mRNA processing
SRRT	Serrate RNA effector molecule homolog	cytoplasm, nucleoplasm	DNA binding	primary miRNA processing, regulation of transcription, transciption, neuronal stem cell maintenance
RCL1	RNA 3'-terminal phosphate cyclase-like protein	nucleolus	catalytic activity	rRNA processing
U3IP2	U3 small nucleolar RNA-interacting protein 2	nucleolus, ribonuleoprotein complex	RNA binding	rRNA processing

depleted in sample

Symbol	Protein name	GO cellular component	GO molecular function	GO biological process
RPL29	26S ribosomal protein L29	ribosome	structural constituent of ribosome	translation
E9Q6E5	Serine/Arginine-rich splicing factor 11	nucleus	nucleotide binding	nd
G3P	Glyceraldehyde-3-phosphate dehydrogenase	cytoplasm, cytoskeleton, nucleus	oxidoreductase, transferase	apoptosis, glycolysis, translation regulation
NOB1	RNA binding protein NOB1	cytoplasm, nucleolus	metal ion binding	maturation of SSU-rRNA
RS4X	40S ribosomal protein S4, X isoform	polysome, small ribosomal subunit	rRNA binding, structural constituent of ribosome	translation
K1C10	Keratin-10	cytoplasm, intermediate filament, keratin filament	structural constituent of epidermis	cellular response to calcium ion, keratinocyte differentiation
RS27	40S ribosomal protein S27	cytosolic small ribosomal subunit	metal ion binding, structural constituent of ribosome	translation
TBB5	Tubulin beta-5 chain	cell body, cytosol, extracellular vesicular exosome, microtubule, nuclear envelope lumen, nucleus, tubulin complex	GTP binding, GTPase activity, structural constituent of the cytoskeleton	protein folding, protein polymerization, spindle assembly

One possible complication of this type of experiments is that it is difficult to distinguish between direct interactors and indirect nucleic acid-mediated interactions. To address this

point the experiment was repeated in the presence of RNase. For unknown reasons the corresponding MS data were of insufficient quality for statistical testing using the power law global error model (Figure 2.10), and this experiment needs to be repeated in the future. Even though the overall overlap and quality of all datasets did not allow for a combined analysis, two of the raw datasets were of sufficient quality for individual evaluation. The raw data clearly indicate that the ATP-dependent RNA helicase A (E9QNN1 also known as DEAH (Asp-Glu-Ala-His) box helicase 9 [Dhx9] or RNA helicase A [RHA]) with sample over control enrichment values of 6 and 9.5 in the first and second IP with RNase, respectively, is the strongest candidate binder.

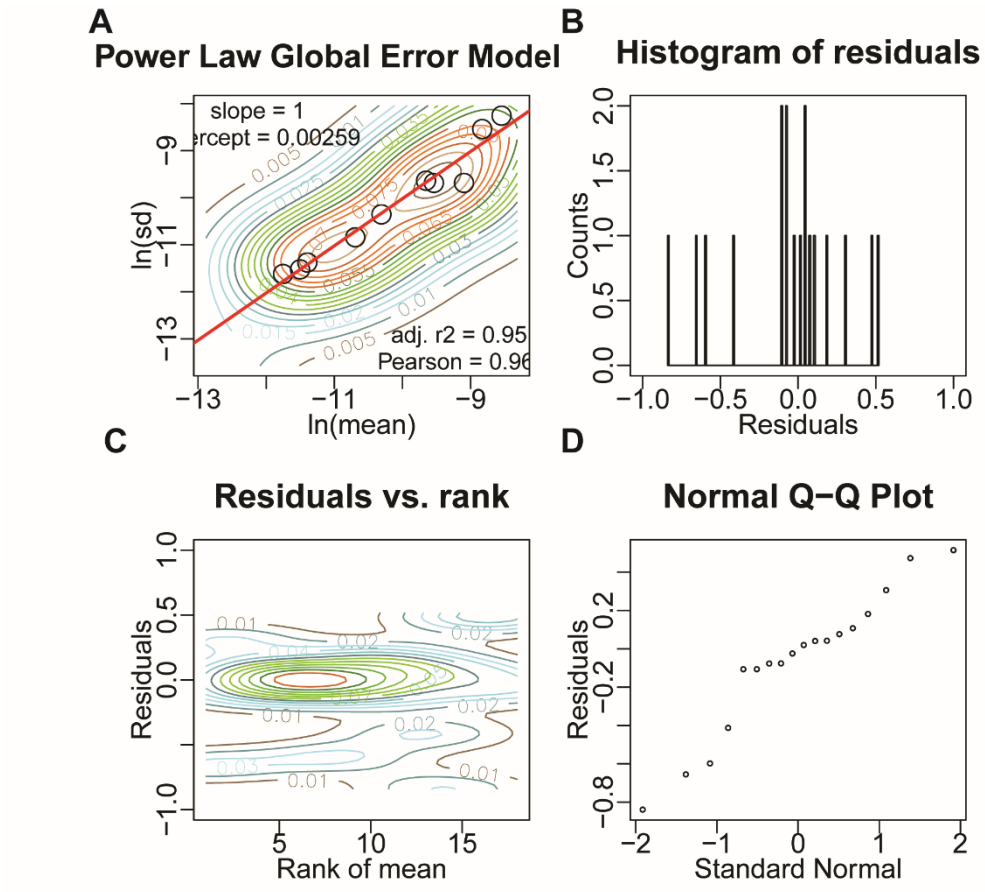


Figure 2.10 – Quality control showing low quality of SINQ IP data in the presence of RNase.

(A) Power law global error model. This plot represents the natural logarithm of the mean (ln(mean)) versus the natural logarithm of the standard deviation (ln(sd)). The low coefficient of determination (R^2) indicates that the model is not suitable for the given set of data. (B)

Histogram of residuals. The histogram presents a plot of residuals versus counts. The number of data points is insufficient to make up a bell-shaped Gaussian distribution curve. (C) Residuals vs. rank. This plot shows the rank of the mean plotted against the residuals. (D) Normal Q-Q plot. Quantile-quantile plot displaying standard normal values plotted against observed residuals. A skew within the dataset is visible.

2.4 SAF-A/hnRNP U is an extensively ubiquitinated protein

Previous studies provided evidence for the presence of post-translational modifications (PTMs) on human SAF-A/hnRNP U. Specifically, SAF-A/hnRNP U was predicted to be a phosphoprotein with several N-linked glycosylation consensus sequences (Kiledjian and Dreyfuss 1992). The observed discrepancy between the predicated and the apparent molecular weight of the protein (~89 kDa and ~120 kDa, respectively) was one of the indicators that lead to this hypothesis. Later studies identified a DNA-dependent protein kinase (DNA-PK)-mediated phosphorylation site at Serine 59 of human SAF-A/hnRNP U (Britton, Froment et al. 2009). Moreover, it was reported that the protein arginine methyltransferase (PRMT1), the predominant PRMT in human cells, binds to and methylates SAF-A/hnRNP U (Herrmann, Bossert et al. 2004).

These observations are especially interesting in the light of earlier studies in which visualization of endogenous SAF-A/hnRNP U-enrichment at the Xi proved to be non-trivial (Helbig and Fackelmayer 2003, Hasegawa, Brockdorff et al. 2010, Pullirsch and Jantsch 2010). One of the possible explanations for why most antibodies fail to highlight the pool of SAF-A/hnRNP U at the Xi, is epitope occlusion caused by PTMs on the protein.

Having good quality MS data available meant that it was possible to run an error tolerant Mascot search. This is a common proteomics tool used for the initial identification of protein post-translational modifications. This approach, however, failed to identify any of the

reported modifications indicating that in the given cellular system SAF-A/hnRNP U is not subject to PTM. Alternatively, it could be that the PTMs are very transient or that the antibody does not pull down modified protein.

Apart from the apparent lack of PTMs the error tolerant search identified a large number of ubiquitination sites marked by residual Glycine-Glycine (Gly-Gly) motifs and incomplete tryptic digestion. Notably, only Lysine (K), Threonine (T), Serine (S) and Cysteine (C) can bear this particular modification (Cadwell and Coscoy 2005, Wang, Herr et al. 2007). All identified sites are listed in Table 2.2.

Table 2.2 – SAF-A/hnRNP U ubiquitination sites identified by error tolerant Mascot searches.

Peptides with candidate ubiquitination sites in red. The location of all identified peptides within the protein primary sequence is displayed in the right column.

Peptide	Modification	Location
K.S**S**GPTSLFAVTVAPPGAR.Q + [+114.0429 at S2]	GlyGly (T)	181-200
S.SGP**T**SLFAVTVAPPGAR.Q + [+114.0429 at T4]	GlyGly (T)	182-200
HLY**T**KDIDIHEVR.I + [+114.0429 at T4]	GlyGly (T)	323-337
K.NGQDLGVAF**K**ISK.E + [+114.0429 at K10]	GlyGly (T)	399-413
K.D**C**EVVMMIGLPGAGK.T + 2 Oxidation (M); [+114.0429 at C2]	GlyGly (T)	471-487
K.QMAD**T**GKLNTLLQR.A + Oxidation (M); [+114.0429 at T5]	GlyGly (T)	520-535
K.QMADTG**K**LNTLLQR.A + Oxidation (M); [+114.0429 at K7]	GlyGly (T)	520-535
R.APQCLG**K**FIEIAAR.K + [+114.0429 at K7]	GlyGly (T)	534-549
R.**K**MCLFAGFQR.K + Oxidation (M); [+114.0429 at N-term K]	GlyGly (T)	567-578

2.5 SAF-A/hnRNP U phosphorylation is undetectable

In addition to analyzing MS data in search of SAF-A/hnRNP U PTMs, a directed approach to reveal SAF-A/hnRNP U phosphorylation was taken. The SDS-PAGE Phos-tag™ (Wako Pure Chemical Industries, Ltd) technology allows to easily distinguish phosphorylated and unphosphorylated proteins (Wako , Kinoshita and Kinoshita-Kikuta 2011). Phos-tags are functional molecules that specifically bind to phosphorylated ions (Figure 2.11A). The binding event results in a gel retardation of bound proteins in an SDS-PAGE gel. Thus, a gel shift can be used as indicator for phosphoproteins (Figure 2.11B). The technology also

allows for subsequent MS analysis (Wako). For the purpose of this thesis work, however, only a simple SDS-PAGE Phos-tag experiment was conducted as follows : Equal amounts of FSPE nuclear extracts were either untreated (UT), treated with an excess of phosphatase inhibitor (+PI) or with shrimp alkaline phosphatase (+P) and run on a specifically prepared Phos-tag acrylamide gel. Afterwards the protein was transferred onto a WB membrane and probed with the in house anti-SAF-A/hnRNP U antibody. The experiment was conducted under varying conditions, viz. with different Phos-tag concentrations, phosphatases and also with phosphatase treatment combined with a phosphatase inhibitor, and an exemplary experimental result is shown in Figure 2.11C. None of the tested conditions showed a gel shift in the phosphatase inhibitor treated sample.

This result is in line with the outcome of the MS data analysis presented in the previous section (Chapter 2.4). Taken together this may imply that either both techniques are not sensitive and specific enough to detect SAF-A/hnRNP U protein phosphorylation or there are no phosphorylation marks on SAF-A/hnRNP U in FSPE cells. It may also indicate that the phosphorylation event on SAF-A/hnRNP U is very transient and the techniques used may not be able to capture the dynamic nature of the process.

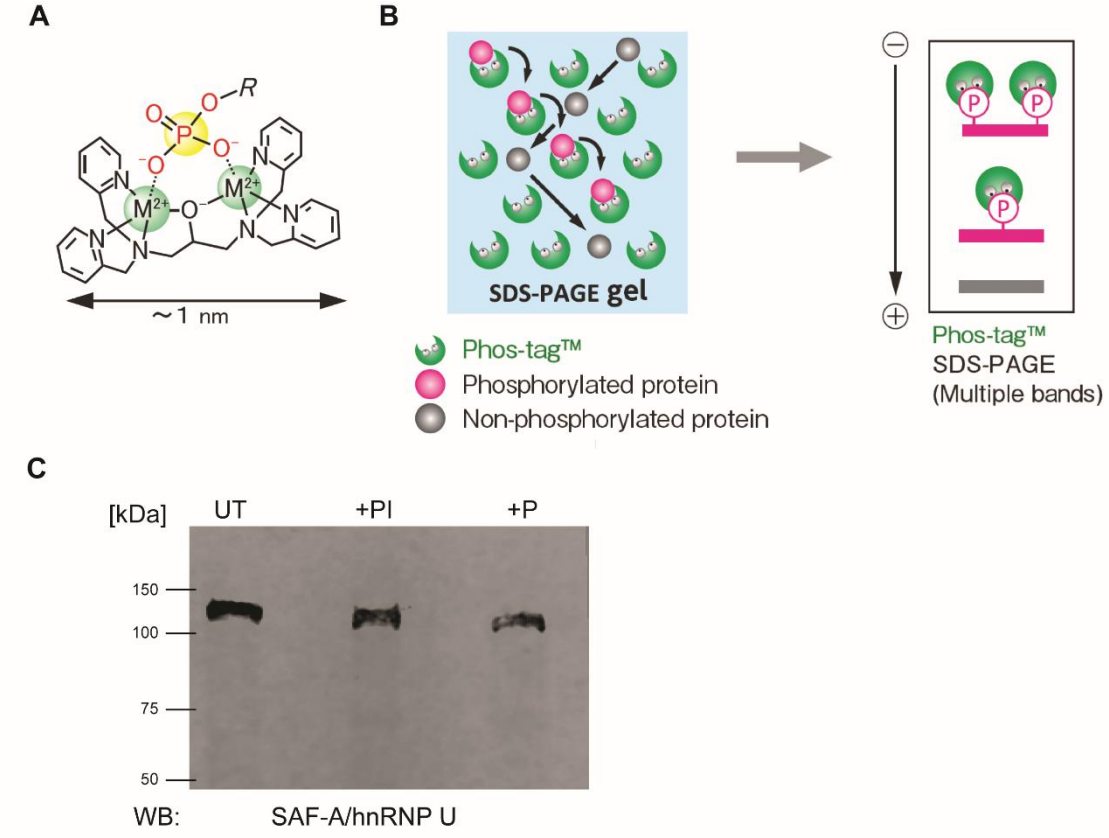


Figure 2.111 – SAF-A/hnRNP U appears unphosphorylated in Phos-tag SDS-PAGE.

(A) Phos-tag basic structure where M^{2+} is either a zinc or manganese ion. The binding selectivity to phosphate ion is much stronger than to any other anion. Therefore, a stable complex is formed under physiological conditions (Wako). (B) Phos-tag principle: Phosphorylated proteins are bound by Phos-tags whilst moving through the SDS-PAGE gel, non-phosphorylated proteins remain unbound. Phos-tag binding decreases migration speed of bound proteins such that phosphorylated and unphosphorylated species are separated (Wako). (C) Anti-SAF-A/hnRNP U WB image of Phos-tag experiments with FSPE nuclear extracts that were either prepared with no additional enzymatic treatment (UT), with addition of a phosphatase inhibitor (+PI) or with a phosphastase (+P). No gel shift was observed in the +PI sample.

2.6 SAF-A/hnRNP U is a nuclear protein with broad nucleoplasmic staining pattern

Early work in human cells (HeLa) showed that the localization pattern of SAF-A/hnRNP U is unrelated to other hnRNPs like hnRNP C, I and L (Mattern, van der Kraan et al. 1999). Furthermore, it was reported that SAF-A/hnRNP U is generally excluded from nucleoli and nuclear speckles. Like hnRNP C, I and L SAF-A/hnRNP U showed a distinctive particulate nuclear staining pattern. Later studies revealed that in female cells SAF-A/hnRNP U seemingly associates with the Xi (Helbig and Fackelmayer 2003, Pullirsch, Hartel et al. 2010). To date the Xi enrichment of endogenous protein had been shown in two instances. First, using a polyclonal antiserum on human cells and secondly, after nuclear scaffold preparations from human embryonic kidney 293 (HEK293) cells, where control stainings, i.e. for H3K27me3 or macroH2A, cannot be used. To investigate SAF-A/hnRNP U staining patterns in mouse cells and see whether the in house reagent shows the enrichment of SAF-A/hnRNP U at the Xi, IF experiments were done on a variety of cell lines and with different fixation conditions.

The first step in the investigation of the antibody's suitability for IF applications was an antibody-antigen competition assay (Figure 2.122). The assay clearly showed that the antigen (0 – 10 µg) was able to compete away the antibody. It also showed a staining pattern that was consistent with the published literature, i.e. a punctuate distribution of the protein and exclusion from nucleoli (Mattern, van der Kraan et al. 1999). In these preliminary IF experiments on a murine female cell line (FSPEs) it became clear that the Xi would not easily be visualized using this reagent.

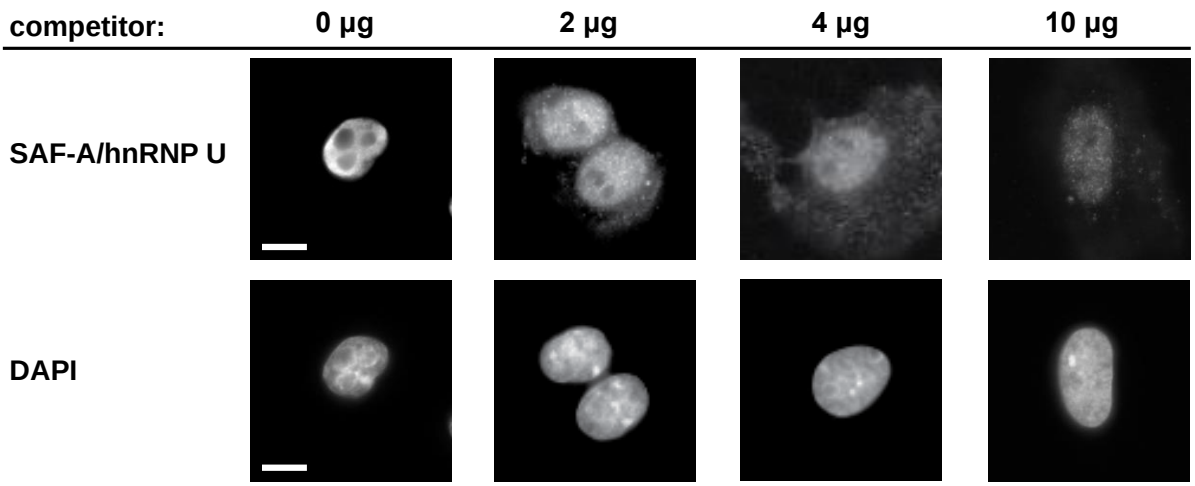


Figure 2.122 – Anti-SAF-A/hnRNP U specifically recognizes the SAF-A/hnRNP U epitope in IF.

IF-based antigen-antibody competition assay on FSPE cells with anti-SAF-A/hnRNP U in the upper panel and 4', 6-diamidino-2-phenylindole (DAPI) staining in the panel below. An increasing amount of SAF-A/hnRNP U antigen (0 µg to 10 µg) was added to the IF preparation. All images were taken using the same exposure time. Scale bar 10 µm.

Therefore, IF imaging under different fixation conditions was tested (Figure 2.133). Again none of the methods showed an improved staining pattern as compared to the standard fixation with formaldehyde, and the expected enrichment at the Xi remained undetectable.

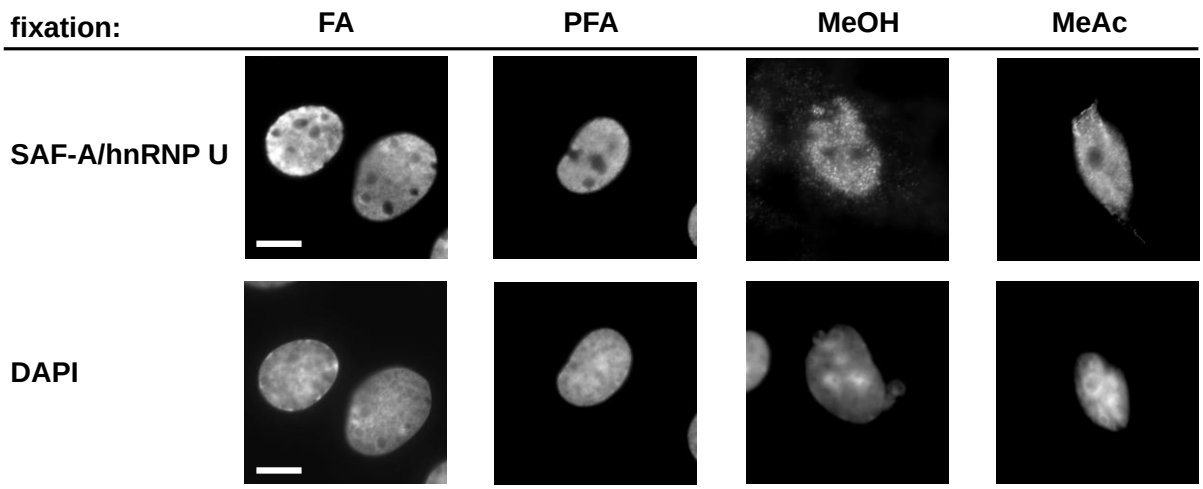


Figure 2.133 – Anti-SAF-A/hnRNP U does not show the enrichment at the Xi using different fixation techniques.

Anti-SAF-A/hnRNP U IF on FSPE cells using different fixation methods; i.e. 2% formaldehyde (FA), 4% paraformaldehyde (PFA), methanol (MeOH) and methanol-acetone (MeAc) fixation (left to right). Nuclear DNA was counterstained with DAPI. Scale bar 10 µm.

Figure 2.14 shows four cell lines, including FSPEs, which were used for testing the antibody's reactivity with SAF-A/hnRNP U. In order to exclude that failure to detect of the Xi related to the cell type used, additional IF experiments on different mouse cell lines were conducted. All lines were co-stained with anti-SAF-A/hnRNP U and anti-H3K27me3. The latter is one of the hallmarks of the Xi and has been used in many studies to pinpoint the inactive chromosome within the nucleus (Hasegawa, Brockdorff et al. 2010, Calabrese, Sun et al. 2012, Escamilla-Del-Arenal, da Rocha et al. 2013). The first cell line tested was Neuro2a. The same murine neuroblast line had been used in a study by Hasegawa *et al.* that provided substantial evidence for the involvement of SAF-A/hnRNP U in XCI (Hasegawa, Brockdorff et al. 2010). Also in Figure 2.14 no Xi-enrichment was visible (compare with Figures 2.12 and 2.13). In contrast, the H3K27me3 antibody clearly highlights the inactive chromosomes in these cells. The next set of cells tested was of embryonic origin (3E ESCs). As mentioned previously, in these cells Xist can be induced by addition of doxycycline to the culturing medium. The image in Figure 2.14 was taken using cells that were induced for 48 hours. Therefore, an accumulation of H3K27me3 was visible at the site of induction (c.f. arrowhead). However, no SAF-A/hnRNP U enrichment at the site of Xist induction was observed. Lastly, trophoblast stem cells (TSCs) that are a good model for imprinted XCI and are known for the good visibility of their Xi in IF, were used to test the reagent (Baumann and De La Fuente 2009). Similarly to the other IF experiments, the inactive chromosomes were clearly visible in the H3K27me3 staining and invisible in the SAF-A/hnRNP U channel.

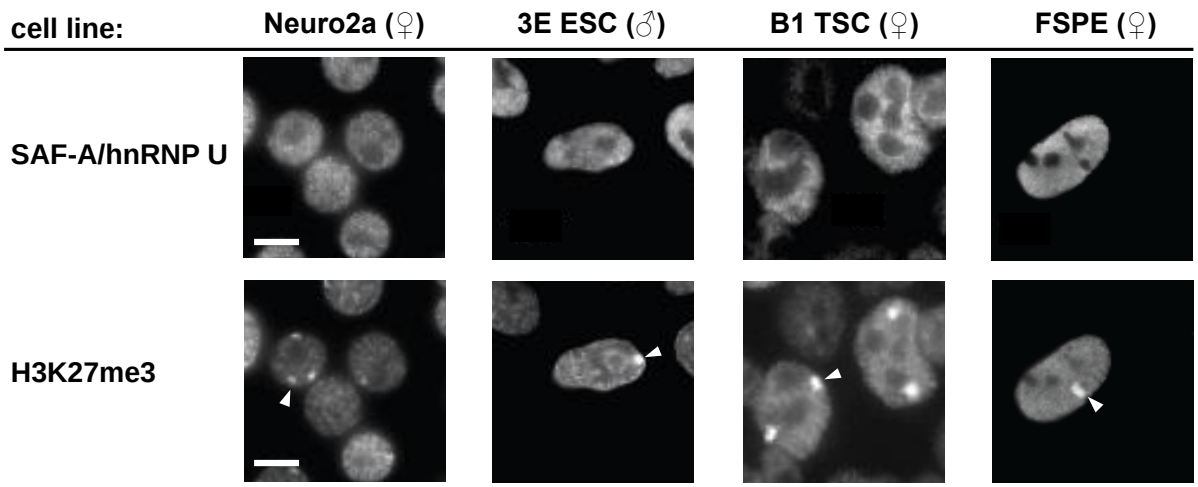


Figure 2.14 – Anti-SAF-A/hnRNP U does not show the enrichment on the Xi in IF on a variety of cell lines.

IF for SAF-A/hnRNP U (top panel) and H3K27me3 (bottom panel) on Neuro2a, 3E ESCs, B1 trophoblast stem cells (TSCs) and FSPEs. All lines - apart from the ESCs that bear an inducible Xist transgene - are female. The arrowheads (▼) point at exemplary H3K27me3 enriched foci. In all lines, apart from the male 3E cells, these correspond to the Xi chromosome(s). Scale bars 10 µm.

In a final attempt to visualize endogenous SAF-A/hnRNP U, a classical *in situ* preparation of the nuclear scaffold was utilized ((Helbig and Fackelmayer 2003) and personal communication with F.O. Fackelmayer). Conventional Xist RNA-FISH showed that Xist is enriched at the Xi in untreated samples (untreated) and scaffold preparations (matrix) of parental FSPEs (c.f. Figure 2.15A and C). However, there was no clear enrichment of endogenous SAF-A/hnRNP U at the Xi when anti-SAF-A/hnRNP U was utilized (Figure 2.15B). On the contrary, a GFP-tagged version of SAF-A/hnRNP U, which is discussed in more detail in Chapter 4, was clearly enriched at the Xi after a matrix preparation (Figure 2.15D). In this case the Xist signal and SAF-A/hnRNP U:hrGFP enrichment were found in the same locus. Notably, hrGFP does not appear to be completely quenched by the formamide used in the RNA-FISH procedure.

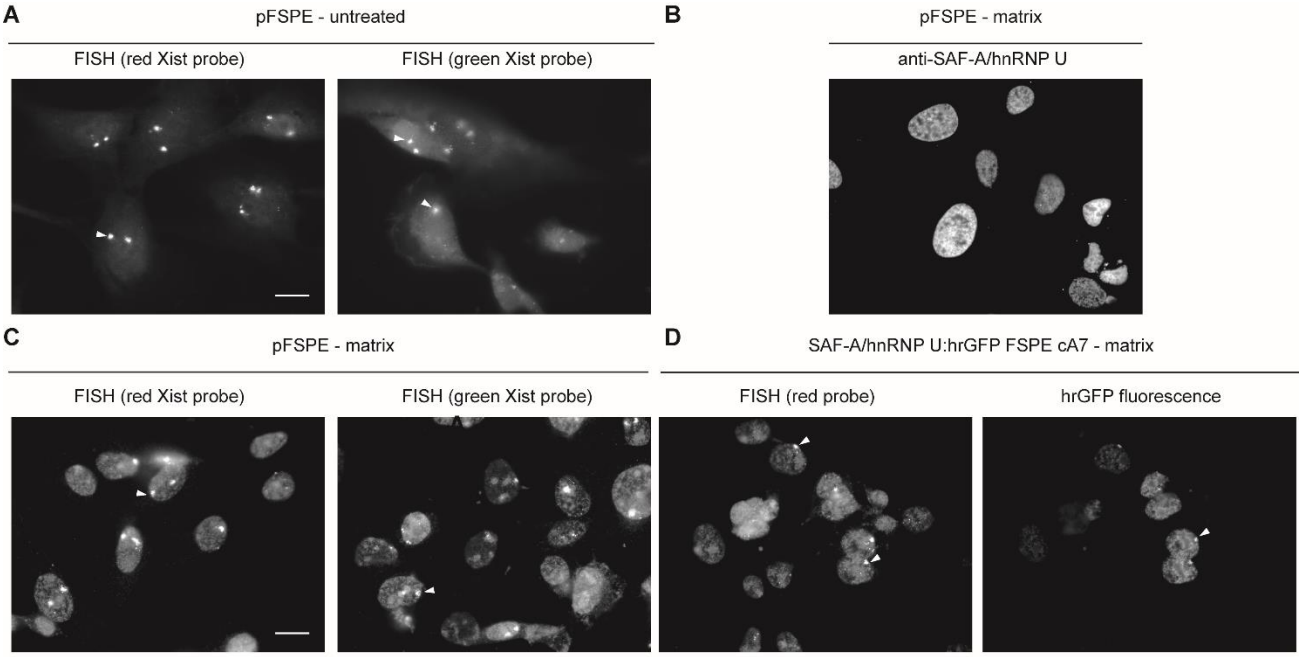


Figure 2.15 – Endogenous SAF-A/hnRNP U enrichment at the Xi cannot be visualized in *in situ* nuclear matrix preparation.

(A) Xist RNA-FISH on parental FSPE cells (pFSPEs) without scaffold extraction (untreated). Red and green FISH probes were utilized. Arrowheads pinpoint individual Xis. (B) Post matrix extraction, anti-SAF-A/hnRNP U immunostaining of pFSPE cells. (C) Xist RNA-FISH with red and green probes on pFSPE nuclear matrices. Arrowheads point at individual, Xist-enriched foci. (D) Matrix extractions of FSPEs stably expressing a SAF-A/hnRNP U transgene. Xist RNA-FISH with red probe (left). hrGFP fluorescence image of the same cells (right). Arrowheads point at foci of Xist and SAF-A/hnRNP U enrichment, respectively. Scale bars 15 µm.

2.7 Summary and discussion

2.7.1 SAF-A/hnRNP U protein interactions

Antibody testing revealed that anti-SAF-A/hnRNP U is suitable for the use in general biochemical and imaging applications. MS analysis showed that the in house reagent is effective in IP applications as it efficiently recognizes SAF-A/hnRNP U. Co-IP experiments suggest that SAF-A/hnRNP U associates with a variety of nucleic acid binders. Thereby the ATP-dependent RNA helicase A E9QNN1 is the strongest interaction candidate. In human

cells this protein is involved in transcription and RNA metabolism (Abdelhaleem, Maltais et al. 2003). Interestingly, DNA-PK phosphorylates E9QNN1 and other hnRNPs (Zhang, Schlott et al. 2004). Other studies associated DNA-PK-mediated SAF-A/hnRNP U phosphorylation with the cellular response to DNA double-strand breaks (Berglund and Clarke 2009, Britton, Froment et al. 2009). These preliminary findings may imply that the two proteins associate to build a functional unit within the cell. The biological significance of the interaction needs to be investigated in the future. Moreover, by using the 3E ESC line that allows for a controlled up-regulation of Xist from an autosomal transgene leaves open the option for follow up experiments in which the SAF-A/hnRNP U interactome would be compared under Xist-induced and non-induced conditions.

2.7.2 SAF-A/hnRNP U post-translational modifications

As has been observed for other hnRNPs, for example hnRNP A1 that is controlled by a set of protein kinases called Mnks, hnRNP U cellular functions may be regulated by PTMs. This study failed to verify previously published evidence on SAF-A/hnRNP U phosphorylation (Britton, Froment et al. 2009). Further attempts to show SAF-A/hnRNP U phosphorylation utilizing Phos-tag gel retardation assays with a reagent that has high affinity to phosphate di-anions failed (Kinoshita, Kinoshita-Kikuta et al. 2012). Instead multiple ubiquitination sites were identified. Ubiquitination is a highly conserved PTM that may regulate protein trafficking, function and turnover (reviewed in Lill and Wertz 2014). Poly-ubiquitination often tags a protein for degradation via the proteasome pathway. No exact function can be assigned to individual sites without looking into the molecular details. Given the rapid turn-over of SAF-A/hnRNP U during the cell cycle (see Chapter 5) it is likely that the majority of ubiquitination events at given sites is involved in protein degradation at specific stages of the cell cycle. There is one other study that reported an association of SAF-A/hnRNP U with the E3 ubiquitin ligase $SCF^{ß\text{-}TrCP}$ (Davis, Hatzubai et al. 2002). However, E3-bound hnRNP U

was found to be stable. Therefore SAF-A/hnRNP U cannot be a true substrate for this particular E3 ligase.

To follow up on the ubiquitination study and link it to SAF-A/hnRNP U turn-over during the cell cycle cells may be synchronized and analyzed by anti-ubiquitin Western blotting and MS.

2.7.3 SAF-A/hnRNP U nuclear localization and visualization at the Xi

Even though this and other studies show that GFP/hrGFP-tagged SAF-A/hnRNP U is enriched at the Xi, visualization of the enrichment of endogenous protein at the Xi with the in house antibody using conventional light microscopy failed (Helbig and Fackelmayer 2003, Pullirsch, Hartel et al. 2010). Nevertheless, Xist RNA-FISH experiments on scaffold preparations of cells stably expressing SAF-A/hnRNP U:hrGFP (Figure 2.15) showed that fluorescently tagged SAF-A/hnRNP U and Xist locate to the same subcellular regions.

There are several possible explanations for the observation that endogenous enrichment of SAF-A/hnRNP U at the Xi is undetectable: One is that a sub-population of SAF-A/hnRNP U is posttranslationally modified (Figure 2.16A). The modified pool is being targeted to the Xi, and consequently masks the antibody epitope(s) through changes in protein conformation and/or direct epitope exclusion. This hypothesis is in line with previous studies reporting on SAF-A/hnRNP U phosphorylation (Berglund and Clarke 2009, Britton, Froment et al. 2009). Since phosphorylation is one of the most common modes of protein PTM that affects protein function, it is likely to affect SAF-A/hnRNP U functionality in one or the other way. Given that the present study was unable confirm this previously published evidence, and that most other studies investigate the human protein, it deems to be the less likely explanation for these observations.

Another possibility is that SAF-A/hnRNP U multimerizes upon DNA nucleotide binding, with the resulting tight packaging of the protein leading to epitope occlusion (Figure 2.16B) (Fackelmayer, Dahm et al. 1994). Furthermore, SAF-A/hnRNP U chains may compact Xi chromatin rendering it a generally inaccessible structure, and/or other proteins may be associated with Xi-bound SAF-A/hnRNP U to form higher-order molecular complexes. The interacting proteins may contribute to the concealment of antibody epitope(s).

Finally, there may be two isoforms of SAF-A/hnRNP U of which only the pan-nuclear isoform is recognized by the antibody used in this study (Figure 2.166C). In fact, there are two other reports supporting the idea that cells produce two forms of SAF-A/hnRNP U. In one of the works two distinct mRNA species of SAF-A/hnRNP U were discovered in human cells (Fackelmayer and Richter 1994). The other study published in the same year, identified two discernible variants of the SAF-A/hnRNP U protein (Fackelmayer and Richter 1994). Notably, these studies have not been followed up on. Further it is unclear whether these two observations are interrelated. A search through the Ensembl database showed that in mouse there may be two protein-coding isoforms of SAF-A/hnRNP U, however, the mRNA of one seems to be subject to nonsense mediated decay. The antibody raised for the purpose of this study was made using the N-terminal third of the SAF-A/hnRNP U protein. Interestingly, there is no variation in this region between the two annotated isoforms. Together with the finding that transgenic SAF-A/hnRNP U:hrGFP, which represents a single protein isoform, is enriched at the Xist RNA-decorated loci in nuclear scaffold preparations this finding disfavors the isoform hypothesis.

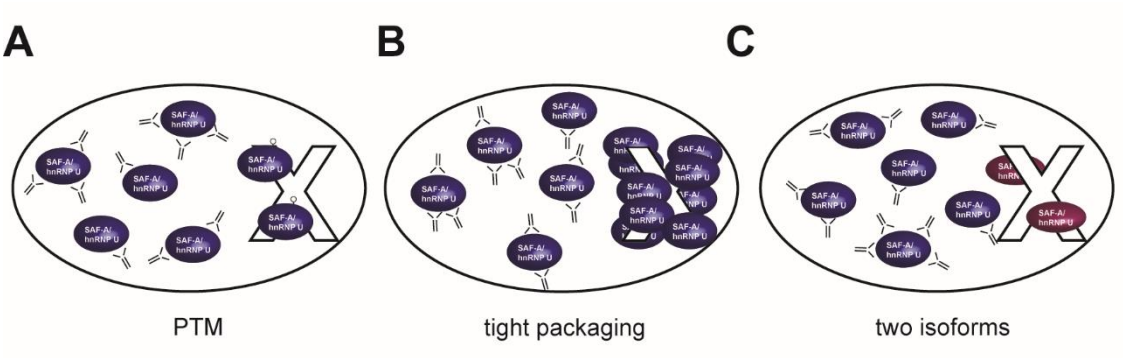


Figure 2.166 – Models for anti-SAF-A/hnRNP U epitope exclusion at the Xi.

Schematic representation of nuclei (ovals) with the Xi (𝕏) and SAF-A/hnRNP U (ellipsoids). (A) Post-translational modification (PTM) model: Post-translationally modified SAF-A/hnRNP U marked by the empty lollipops (Ρ) locates to the Xi. The PTM may induce conformational changes and/or directly mask the anti-SAF-A/hnRNP U (Y) epitope. (B) Tight packaging model: Xi-bound SAF-A/hnRNP U is packed so tightly that the epitope is rendered inaccessible to the antibody. (C) Two isoform model: There may be two different SAF-A/hnRNP U isoforms (blue vs. magenta ellipsoids). Only one of which – unbound SAF-A/hnRNP U - is recognized by anti-SAF-A/hnRNP U.

3. *In vivo* UV crosslinking and immunoprecipitation to characterize SAF-A/hnRNP U RNA binding activity

3.1 Introduction

Early work on SAF-A/hnRNP U indicated that the protein binds to RNA (Dreyfuss, Choi et al. 1984, Kiledjian and Dreyfuss 1992). These studies showed that the protein binds pre-mRNA *in vivo* and RNA *in vitro*. Interestingly, SAF-A/hnRNP U does not contain the canonical RNA binding domain consensus sequence that is found in many other hnRNPs (Adam, Nakagawa et al. 1986, Burd, Swanson et al. 1989). Hence, the protein's RNA binding region had to be identified by deletion studies. Thereby it was discovered that the C-terminal RGG domain is sufficient for SAF-A/hnRNP U to bind RNA (Figure 3.1). Owing to its amino acid composition the RGG box is very basic. But it is uncertain how RNA is bound by the domain, and how many RG-rich repeats are required for RNA binding.

Initial binding studies were performed using simple homopolymeric RNA target sequences. Thereby, poly(G) RNA proved to bind strongest, poly(A) and poly(U) sequences had intermediate binding affinity, and poly(C) RNA was bound only weakly (Kiledjian and Dreyfuss 1992). A later study showed that SAF-A/hnRNP U interacts with DNA Topoisomerase IIß through a subset of polyadenylated RNAs, thus, regulating gene expression (Kawano, Miyaji et al. 2010). This finding is supported by another report suggesting that SAF-A/hnRNP U enhances gene expression through mRNA stabilization (Yugami, Kabe et al. 2007). Later Hasegawa *et al.* (2010) provided evidence for the association of SAF-A/hnRNP U with Xist RNA (Hasegawa, Brockdorff et al. 2010). More specifically, they showed that RGG domain deletions strongly affected Xist chromosomal localization, and suggested that SAF-A/hnRNP U interacts with Xist RNA exon 1 (nucleotides 2339– 2525 and 5125 – 5227). *In silico* predictions support these findings and put forward the idea that SAF-A/hnRNP U also interacts with RepA of Xist (Agostini, Cirillo et

al. 2013). However, the observations made by Hasegawa *et al.* remain to be unambiguously reproduced, especially as the techniques used are known to be prone to artifacts. This applies particularly to crude RNA IPs with a quantitative real-time polymerase chain reaction (qRT-PCR) output.

```
                10        20        30        40        50        60        70        80
MSSSPVNVKKLKVSELKEELKKRRLSDKGLKADLMDRLQAALDNEAGGRPAMEPGNGSLDLGGDAAGRSGAGLEQEAAAG

                90       100       110       120       130       140       150       160
AEDDEEEEGIAALDGDQMELGEENGAAGAADAGAMEEEEAASEDENGDDQGFQEGEDELGDEEEGAGDENGHGEQQSQPP

               170       180       190       200       210       220       230       240
AAAAQQQPSQQRGAGKEAAGKSSGPTSLFAVTVAPPGARQGQQQAGGDGKTEQKGGDKKRGVKRPREDHGRGYFEYIEEN

               250       260       270       280       290       300       310       320
KYSRAKSPQPPVEEEDEHFDDTVVCLDTYNCDLHFKISRDRLSASSLTMESFAFLWAGGRASYGVSKGKVCFEMKVTEKI

               330       340       350       360       370       380       390       400
PVRHLYTKDIDIHEVRIGWSLTTSGMLLGEEEFSYGYSLKGIKTCNCETEDYGEKFDENDVITCFANFETDEVELSYAKN

               410       420       430       440       450       460       470       480
GQDLGVAFKISKEVLADRPLFPHVLCHNCAVEFNFGQKEKPYFPIPEDCTFIQNVPLEDRVRGPKGPEEKKDCEVVMMIG

               490       500       510       520       530       540       550       560
LPGAGKTTWVTKHAAENPGKYNILGTNTIMDKMMVAGFKKQMADTGKLNTLLQRAPQCLGKFIEIAARKKRNFILDQTNV

               570       580       590       600       610       620       630       640
SAAAQRRKMCLFAGFQRKAVVVCPKDEDYKQRTQKKAEVEGKDLPEHAVLKMKGNFTLPEVAECFDEITYVELQKEEAQK

               650       660       670       680       690       700       710       720
LLEQYKEESKKALPPEKKQNTGSKKSNKNKSGKNQFNRGGGHRGRGGFNMRGGNFRGGAPGNRGGYNRRGNMPQRGGGGG

               730       740       750       760       770       780       790       800
SGGIGYPYPRGPVFPGRGGYSNRGNYNRGGMPNRGNYNQNFRGRGNNRGYKNQSQGYNQWQQGQFWGQKPWSQHYHQGYY
```

Figure 3.1 – Primary amino acid sequence of murine SAF-A/hnRNP U.

Residues 1 to 800 with arginine (R) and glycine (G) residues in red. These are clearly enriched in the C-terminal proportion of the protein.

These observations prompted the re-investigation of SAF-A/hnRNP U's RNA binding behavior using the state of the art technique *in vivo* UV crosslinking and immunoprecipitation followed by high-throughput sequencing (CLIP-Seq) (Figure 3.2) (Ule, Jensen et al. 2005).

CLIP-Seq utilizes 254 nm ultraviolet (UV) light to covalently link proteins to nucleic acids. UV light is said to be a zero-length crosslinking agent, and several photochemical mechanisms have been reported for protein nucleic acid crosslinking (Williams and Konigsberg 1991, reviewed in Meisenheimer and Koch 1997, Zhang, Zhang et al. 2004, Sei and Conrad 2014). Briefly, the proposed mechanisms commonly assume an initial radical formation step upon photoexcitation of either a nucleic or an amino acid. The free radicals then initiate covalent bond formation. This event is followed by hydrogen abstraction from an adjacent amino acid and results in chemical bond formation. Once proteins and nucleic acids have been crosslinked, the cells are lysed and cellular RNA is digested into 40 to 80 nucleotide fragments and DNA is removed using DNase treatment. This is followed by an IP of the protein of interest that can be carried out under high-stringency conditions and significantly reduces the background signal as compared to other methods. Then a radioactive end-label is added to the RNA in isolated complexes. Subsequently, the samples are further purified by running the RNA-protein adducts on a SDS-PAGE gel from where they are blotted onto a nitrocellulose membrane. At this point the protein-RNA complexes of interest can be visualized by autoradiography and a band at the corresponding running distance is excised off the membrane. From there they are eluted, and all protein is digested away to facilitate the reverse transcriptase reaction that is required for the preparation of a complementary DNA (cDNA) library. Once a library has been obtained it can be sequenced and analyzed bioinformatically upon return of the dataset (for a comprehensive description of the protocol see Wang, Tollervey et al. 2009).

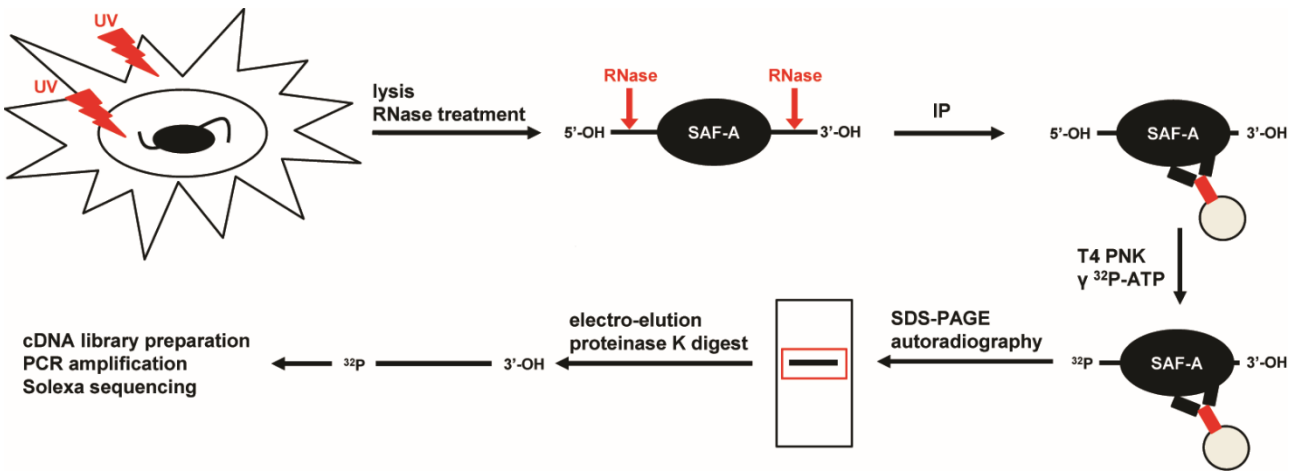


Figure 3.2 – Workflow for an *in vivo* UV crosslinking and immunoprecipitation high-throughput sequencing experiment (CLIP-Seq).

Initially, protein is UV crosslinked to nucleic acids. This step is followed by cell lysis, RNase treatment, immunoprecipitation (IP), end-labeling with γ^{32} P-adenosine triphosphate (ATP), SDS-PAGE and autoradiography. The protein-RNA complexes of interest are recovered by electroelution. After proteolytic digest a complementary DNA (cDNA) library is prepared. Subsequently the library is sequenced.

Data obtained by CLIP-Seq can provide information on which RNAs are bound by a protein under physiological conditions. Furthermore, RNA consensus sequences can be identified (Rinn and Ule 2014). To verify SAF-A/hnRNP U binding to Xist RNA and to identify binding consensus motif(s), a CLIP-Seq experiment using nuclear extracts from 24 hours Xist-induced mouse ESCs and the in house anti-SAF-A/hnRNP U antibody was set up.

3.2 SAF-A/hnRNP U CLIP - Preparatory experiments

Before performing the full length experiment, a few technical issues had to be addressed. The ESC line used had been prepared and characterized by previous laboratory members (refer to Brockdorff lab, Tang, Huntley et al. 2010, Cerase, Smeets et al. 2014). Briefly, a transgenic male mouse ESC line was derived. The cells of this line bear a doxycycline-inducible Xist-transgene on chromosome 17 and will be referred to by its clone number (3E)

in the further discussion (Figure 3.3). Previous studies showed that 24 hours induction with doxycycline is sufficient to effectively up-regulate Xist RNA expression (Cerase, Smeets et al. 2014). That makes the cell line a good tool for all applications where regulated, strong induction of Xist RNA is required. Thence, to initially assess whether SAF-A/hnRNP U-Xist RNA associations can be identified by CLIP, 24 hour Xist-induced 3E cells were used for the IP input preparation.

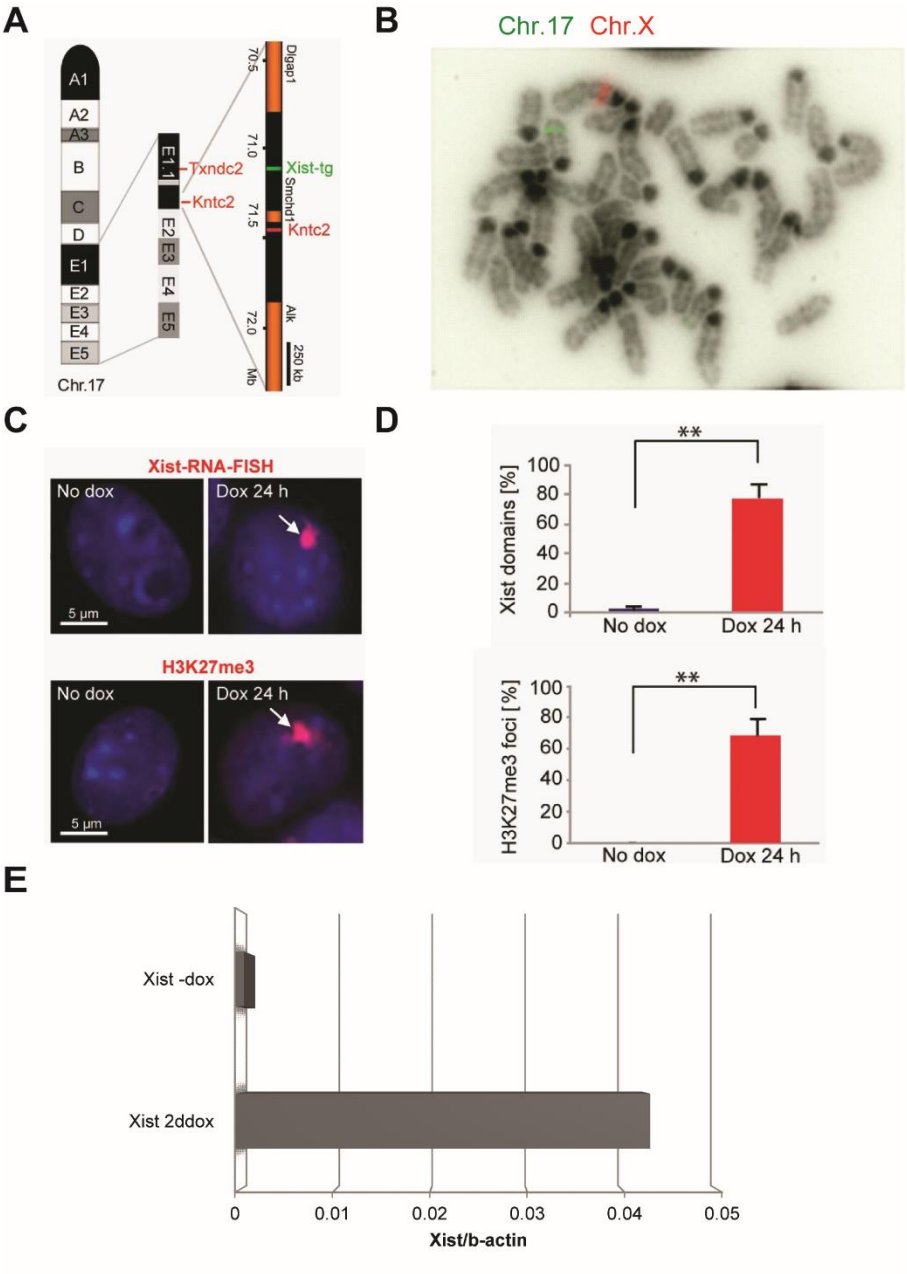


Figure 3.3 – Characterization of a male 3E ESC line carrying a doxycycline-inducible autosomal Xist transgene.

(A) Schematic of chromosome 17 (Chr.17) bearing the doxycycline-inducible Xist transgene (Xist-Tg, green). (B) DNA *in situ* fluorescence hybridization (FISH) on metaphase spreads mapping the Xist-Tg to chromosome 17 and the X chromosome using dioxygenin-11-dUTP or biotin-16-dUTP labeled plasmid DNA as probes (Roche). Transgenic Xist was detected with pTREXist, and endogenous Xist was detected with a BAC clone covering about 100 kb of genomic DNA surrounding and enclosing the Xist locus (by courtesy of Dr. Andrea Cerase) (Tang, Huntley et al. 2010). (C) Fluorescence microscopy images of representative Xist-

induced and uninduced (Dox 24 h and No dox, respectively) 3E ESCs. Xist RNA-FISH and anti-H3K27me3 staining show that after 24 hours of induction, transgenic Xist RNA is expressed and silencing marks are set (by courtesy of Dr. Andrea Cerase) (Cerase, Smeets et al. 2014). (D) Quantification data for Xist RNA and H3K27me3 foci (*n* ≥ 200, error bars correspond to the SD of three independent experiments, two asterisks (**) indicate high statistical significance (*P* ≤ 0.001, *t* test) (by courtesy of Dr. Andrea Cerase) (Cerase, Smeets et al. 2014). (E) qRT-PCR quantification of Xist expression after 48 hours of doxycycline induction (Xist 2ddox) compared with no induction (Xist –dox) normalized to ß-actin.

Previous reports showed that, at least during interphase, Xist RNA mainly associates with the insoluble nuclear matrix fraction, and it may only be released upon entry into or progression through mitosis (Clemson, McNeil et al. 1996, Duthie, Nesterova et al. 1999). Therefore, two parameters had to be tested before carrying out an actual CLIP experiment: First of all the amount of SAF-A/hnRNP U protein found in nuclear soluble vs. pellet fraction was assessed (Figure 3.4). When performing Western blot analysis on equal quantities of protein in the various fractions of a nuclear extract preparation it was found that about half of the total nuclear SAF-A/hnRNP U is in the soluble fraction. This means that most of the nuclear SAF-A/hnRNP U is easily accessible. At this point it is worth mentioning that a sonication step could potentially increase the proportion of soluble SAF-A.

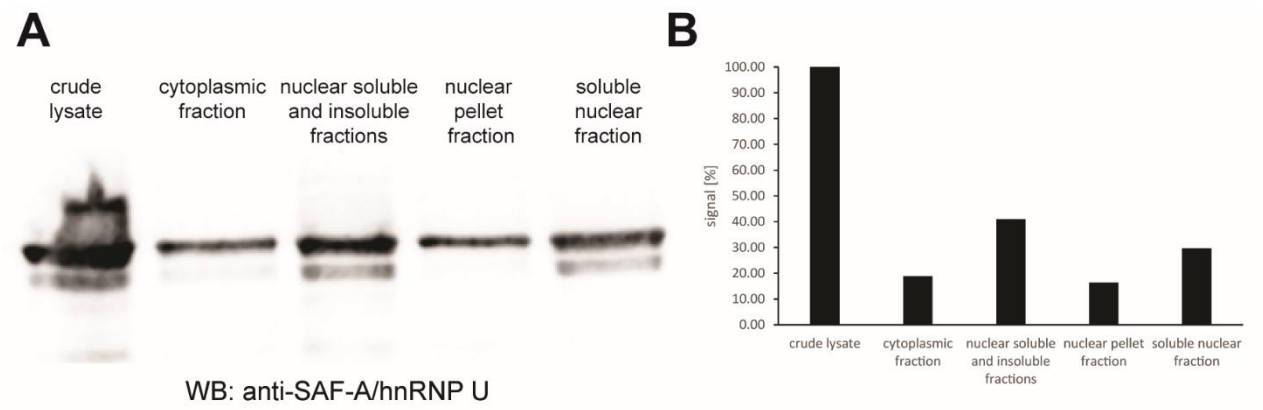


Figure 3.4 – SAF-A/hnRNP U protein in nuclear extract preparation fractions.

(A) Comparative Western blot (WB) analysis of different nuclear extract fractions. SAF-A/hnRNP U levels in crude lysate, cytoplasmic, combined nuclear soluble and insoluble, nuclear pellet and soluble nuclear fractions are shown here. (Note that no loading control is shown as any other cellular protein is expected to be present at varying amounts in the different fractions as well.) (B) Quantification of Western blot signal in the different fractions showing that there is almost twice as much SAF-A/hnRNP U in the soluble nuclear fraction as compared to the nuclear pellet fraction.

Secondly, Xist RNA partitioning within the nuclear fractions needed to be determined as well. This was done by extracting RNA from the various fractions and assessing their Xist RNA content by qRT-PCR with Xist-specific primers (Figure 3.5). Additionally, it was tested whether UV treatment or sonication affect solubilization of Xist RNA. Overall no major differences between plus and minus sonication conditions were seen. Surprisingly, UV treatment appeared to increase Xist extractability.

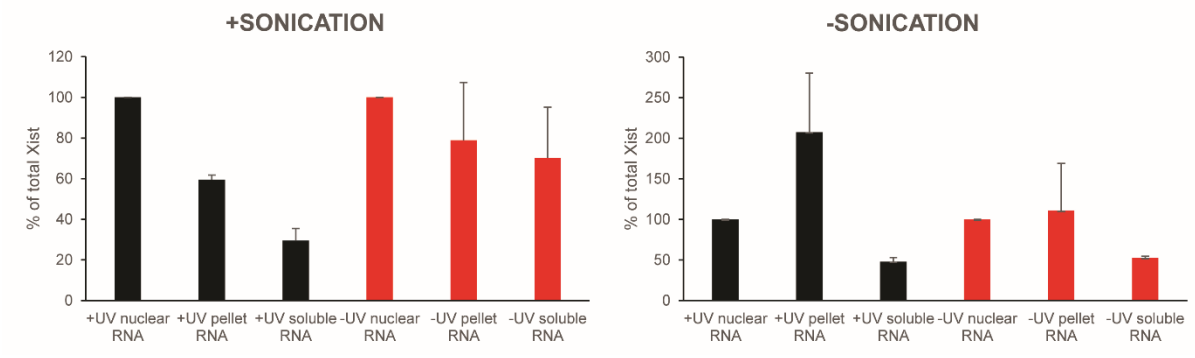


Figure 3.5 – Xist content of nuclear extract preparations.

Quantification of Xist RNA by qRT-PCR. UV treatment (+UV, black bars) is compared with no treatment (–UV, red bars). Furthermore, the effect of sonication on Xist RNA solubilization was investigated. The bar chart on the left-hand side shows normalized Xist levels in the different fractions relative to total Xist RNA in whole nuclei after sonication (+SONICATION). The bar chart on the right shows the same fractions but without sonication (-SONICATION). Notably, the nuclear RNA fraction comprises both pellet and soluble RNA fractions (whole nuclear fraction). Error bars represent the standard error of three independent experiments that is the

STDEV/SQRT(n) where STDEV is the standard deviation and SQRT(n) the square root of the number of experiments. The low amount of +UV nuclear RNA as compared to +UV pellet RNA in the –SONICATION test is probably due to a technical error.

Later, preliminary CLIP experiments showed that sonication significantly increases the non-specific background signal up to a point at which the experiment cannot be conducted (data not shown). It was therefore decided to use unsonicated Xist-induced mouse ESC nuclear extract as input for the CLIP experiment.

3.3 SAF-A/hnRNP U CLIP experiment

After optimizing multiple experimental parameters, a full CLIP experiment was conducted (for a schematic overview see Figure 3.2). The first quality control point in a CLIP experiment is the autoradiography image taken just after the IP step has been completed (Figure 3.6). In the IgG control sample a background signal was observed under UV positive conditions, whereas there was none when cells used for the input were not irradiated with UV light. The UV positive SAF-A/hnRNP U sample shows a clear band with a higher amount of RNase. This band corresponds to SAF-A/hnRNP U-RNA complexes and is an indicator for where to cut a band for RNA extraction and cDNA library preparation. The adjacent low RNase sample is the one of actual interest because the sequencing sample corresponds to the RNA smear just above the molecular weight of mouse SAF-A/hnRNP U. It was roughly estimated that a 20 kDa shift in molecular weight corresponds to protein bound to RNA of an average size of 65 kDa, which was the minimum length required to obtain meaningful results (Wang, Tollervey et al. 2009).

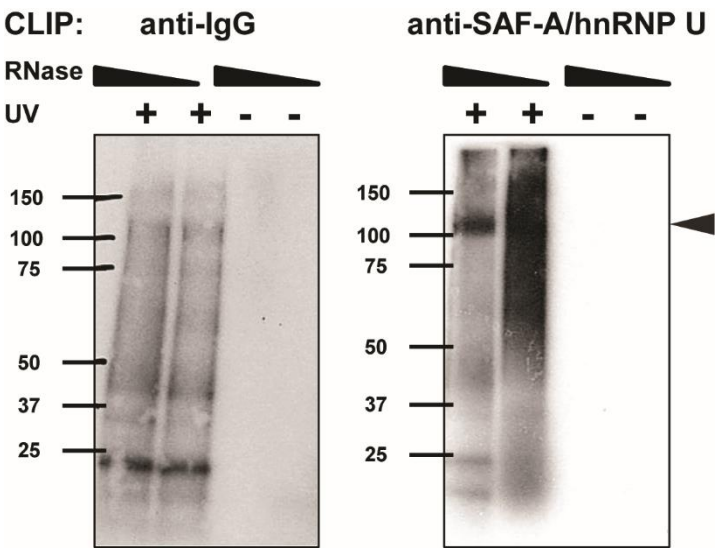


Figure 3.6 – SAF-A/hnRNP U CLIP autoradiography.

First quality control step in a CLIP experiment showing an image of the autoradiography film for the anti-IgG (control) and the anti-SAF-A/hnRNP U (sample) IP at the same exposure time. IPs were conducted on nuclear extracts from cells exposed to 100 mJ/cm^2 (+) and 0 mJ/cm^2 (-) of 254 nm UV light that were treated with two different amounts of RNase I. The arrowhead points at the molecular weight of SAF-A/hnRNP U. A band is visible in the high RNase, +UV anti-SAF-A/hnRNP U IP.

Once the RNA had been extracted, a cDNA library was prepared. For quality control purposes the library was run on a Bioanalyzer (Figure 3.7A). Subsequently, a library aliquot was used for TOPO TA cloning (Fontes 2013). 30 individual bacterial colonies were picked, grown up, plasmids were extracted and Sanger sequenced (Figure 3.7B). The obtained sequences were analyzed using the National Center for Biotechnology Information Basic Local Alignment Search Tool (NCBI BLAST). A majority (89%) of sequences originated from mouse RNA and only a few were purely adapter sequences or sequences too short for proper alignment to a genome. Noteworthy, among the TOPO-cloned mouse sequences were also duplicates, triplicates and even quadruplicates. This may be due to over-amplification of the library in the cDNA preparation step, which is one of the commonly encountered problems in cDNA preparations from very low amounts of input RNA (Mao,

Sendler et al. 2014). But it is equally likely that SAF-A/hnRNP U only binds to a small subset of nuclear RNAs. Therefore it was decided to sequence the present library and perform a preliminary analysis that could be used for the design and optimization of future experiments.

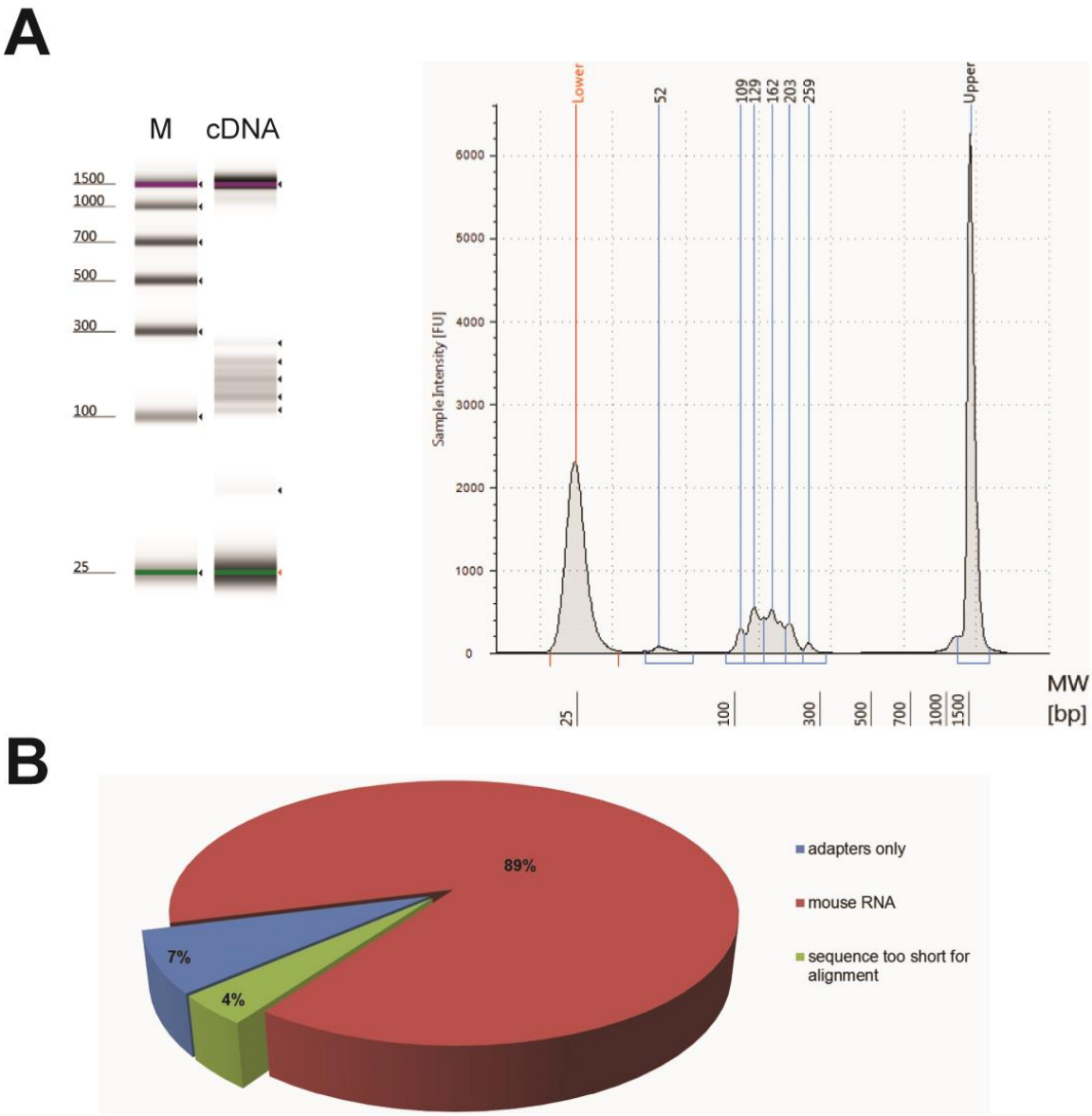


Figure 3.7 – SAF-A/hnRNP U CLIP cDNA library preparation and initial quality control.

(A) Bioanalyzer profile of the cDNA library (cDNA) and a base pair marker (M). Gel bands on the left and a histogram profile with peaks assigned where indicated by arrowheads (▼) on the gel. (B) Result of the TOPO-clone pre-screening. 30 clones were analyzed after bacterial

colony picking and conventional sequencing. The obtained sequences were submitted to a Basic Local Alignment Search Tool (BLAST) search. The outcome indicated that most of the library (89%) was derived from mouse RNA. All remaining sequences were either adapters only or too short to be aligned. Within the 89% a few reads were present in duplicates, triplicates and even quadruplicates. This may either indicate that naturally SAF-A/hnRNP U has a restricted number of RNA target sequences or that the library had been slightly over amplified. The latter reasoning is one of the known pitfalls of CLIP experiments where minute quantities of RNA are amplified to make a cDNA library. Further attempts were made to obtain better quality libraries but all of these failed. Therefore, the present library was sent for high-throughput sequencing (single end, 100 nucleotides, Illumina, performed by the WTCHG).

3.4 SAF-A/hnRNP U CLIP-Seq data quality control

Upon receipt of the raw data, reads were aligned to the mouse genome (mm10) (Table 3.1). For comparison an RNA input dataset from E14 ESC was included in the analysis. The total number of reads in all datasets were on a similar scale. Of all three sets the CLIP-Seq data comprised the highest number of reads indicating great sequencing depth. However, the percentage of mapped reads in this particular sample was lower than in the controls. This could be due to the fact that the control data map fairly evenly across the whole genome, whereas the CLIP data have a restricted set of reads that cover only specific parts of the genome. Therefore, there may be some reads that simply did not map, such as repeat sequences that cannot be mapped unambiguously. Apart from the percentage of mapped reads all remaining parameters, coverage and Picard duplication rate, were in approximately the same range.

Table 3.1 – CLIP-Seq RNA read alignment overview.

The SAF-A/hnRNP U CLIP library was aligned to the mouse genome (version mm10) together with two replicates of whole mouse ESC RNA (E14 control rep1 and E14 control rep 2) libraries. The table compares the number of total and mapped reads within each library. Not all reads of the CLIP library were mappable. The coverage was determined using the formula: coverage = (mapped reads x read length) / total number of bases in genome/transcriptome. All Picard duplication rates are in a similar range.

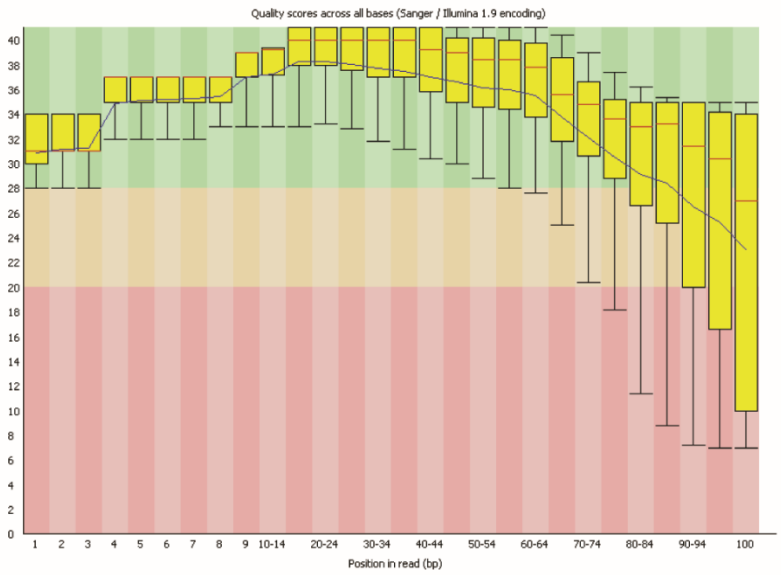
	Release / Version	Total bases	Read length
Genome	mm10	2,730,871,774	100

	SAF-A/hnRNP U CLIP	E14 control rep 1	E14 control rep 2
Total reads	390,012,694	218,393,198	286,952,325
Mapped	257,664,962	218,393,198	286,952,325
Mapped [%]	66.07	100.00	100.00
Coverage	9.44	8.00	10.51
Picard duplication rate	0.43	0.44	0.49

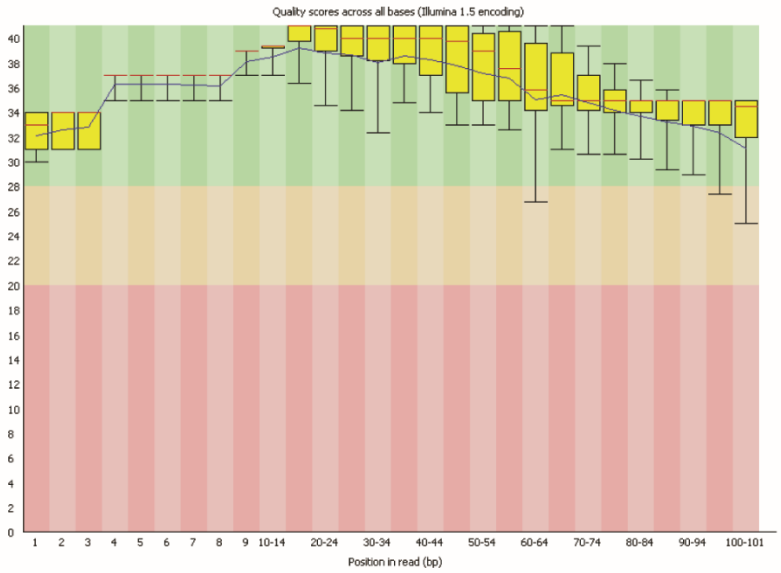
Before drawing any biological conclusions from the data some basic quality controls were performed. For this purpose the FastQC bioinformatics tool kit developed by the Babraham Institute (http://www.bioinformatics.babraham.ac.uk/projects/fastqc) was utilized. Figures 3.8 to 3.14, and Table 3.2 summarize the FasQC results.

Firstly, the per base pair quality of the libraries was evaluated (Figure 3.8). The information derived from this quality control parameter can be used to determine whether and how far a given library needs to be trimmed for the further analysis. Owing to the length of the sequencing runs (100 nucleotides) and the general Chemistry of a sequencing reaction a general trend to quality degradation can be observed towards the end of the sequencing tags (Andrews).

SAF-A/hnRNP U library



Control library 1



Control library 2

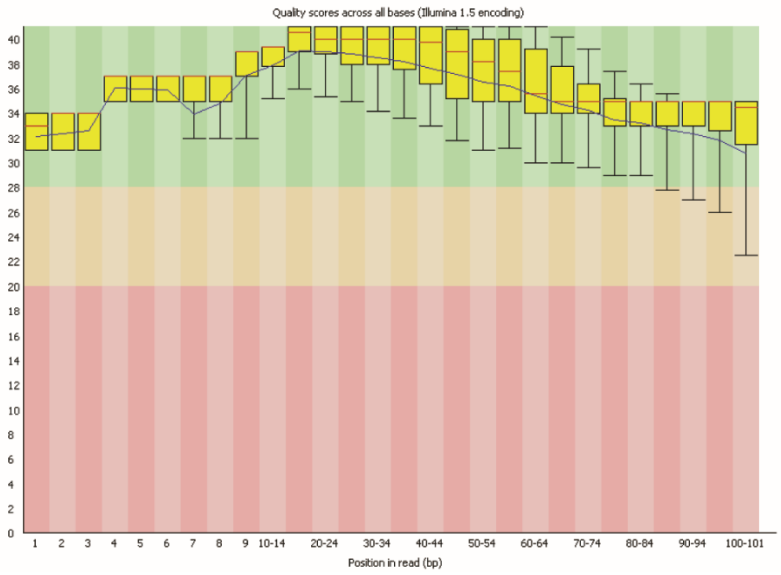
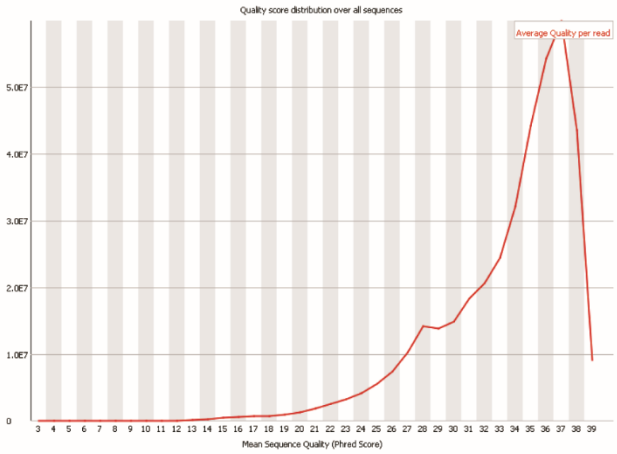


Figure 3.8 – Per base pair sequence quality of SAF-A/hnRNP U CLIP sample and E14 input controls (libraries 1 and 2).

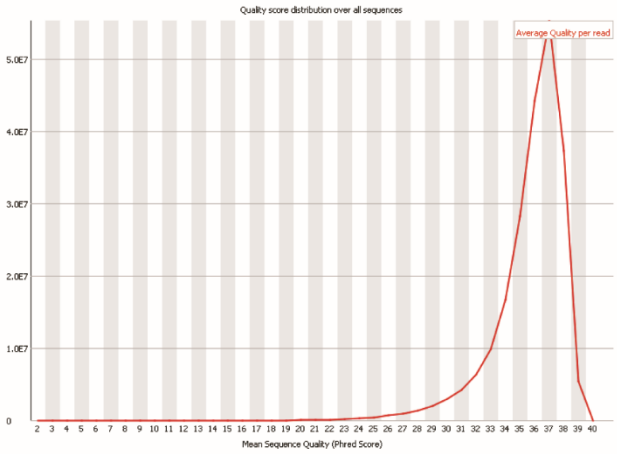
The x-axis shows all of the individual bases for the sequencing reads. For every base a quality core had been assigned (y-axis). The yellow boxes show the interquartile range from the 25th to the 75th quartile, the red line indicates the position of the median, the black whiskers go between the 10th and the 90th percentile and the blue line is the mean quality score. In an ideal sample the quality score stays high across the entire sequencing run. Quality scores in the control libraries stay relatively high up to the end. For the sample the quality is poor starting from position 80-84. This indicates that for more in depth analysis this library would need to be trimmed back to about base 80.

Next, the per sequence quality of the three libraries was inspected (Figure 3.9). This parameter can identify biases/loss of quality in certain subsets of the data but is usually quite robust and trimming may solve issues arising at this point (Andrews). As shown in Figure 3.9, the per sequence quality of the control library is good. In comparison, the SAF-A/hnRNP U CLIP-Seq library has a slight shoulder to the left that may hint at a systematic problem, e.g. in the read out of the flow cell. Notably, the control and the sample libraries were neither prepared in parallel nor sequenced together which is why it is not possible to get to the bottom of this issue at this point.

SAF-A/hnRNP U library



Control library 1



Control library 2

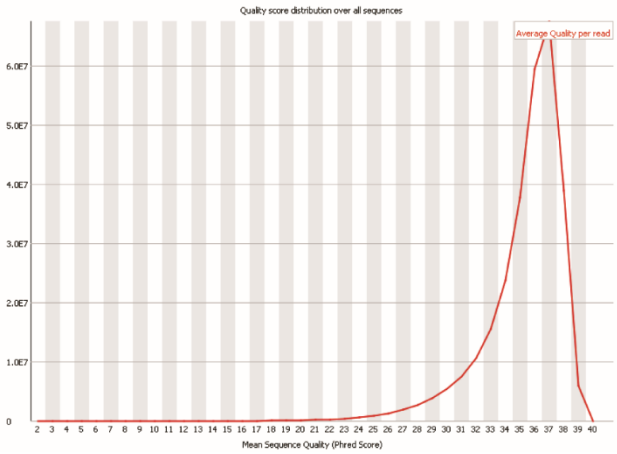
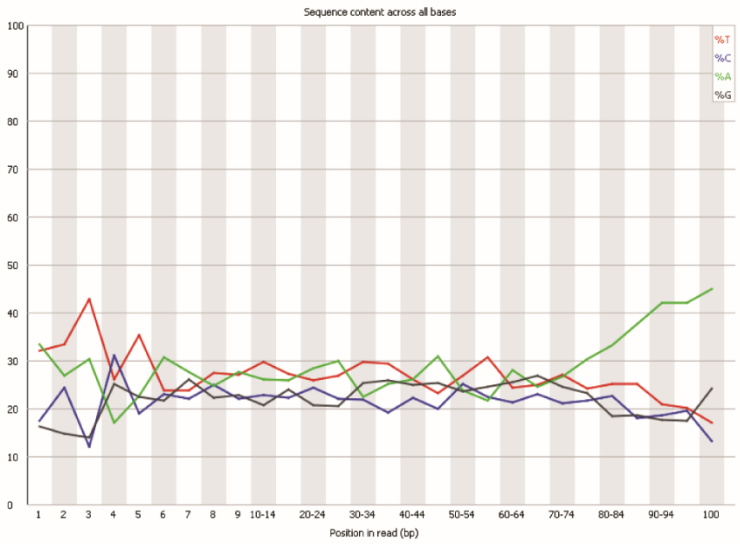


Figure 3.9 – Per sequence quality of SAF-A/hnRNP U CLIP sample and E14 input controls (libraries 1 and 2).

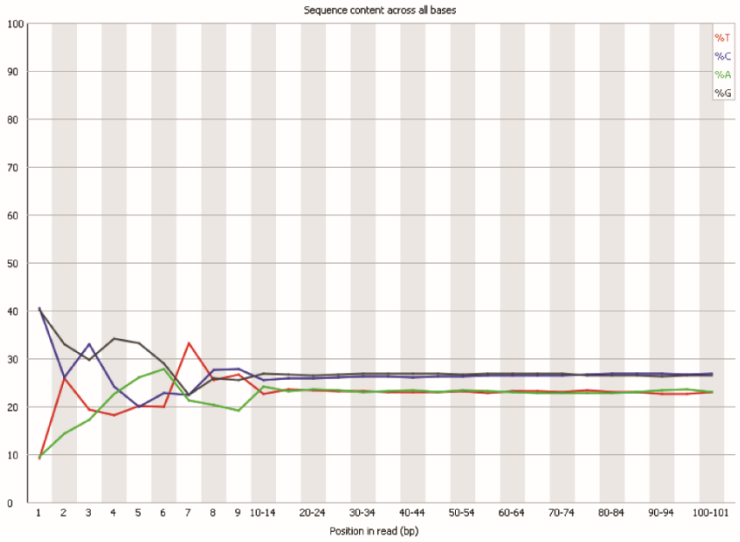
Mean score of sequences across all bases. The peak towards the right hand side of all graphs indicates that in all three libraries all sequences are of reasonable quality. The broader shoulder to the left had side of the SAF-A/hnRNP U library plot indicates that this library is of slightly poorer quality than the two input controls.

In the following the per base sequence content was assessed (Figure 3.10). This representation shows the proportion of each of the four normal bases at different positions in the sequencing tag sets (Andrews). In an ideal, randomized setting it would be expected to have no or little variation between the bases at different positions of the sequencing run which would result in a representation of parallel, superimposed lines. Figure 3.10 shows that after some initial variability had been overcome in the two control libraries the four lines are almost parallel. In the SAF-A/hnRNP U library, however, there is quite some variation between the base pair distribution at the different positions. Such variability may have multiple origins. But in this case it is more likely that certain sequences are overrepresented.

SAF-A/hnRNP U library



Control library 1



Control library 2

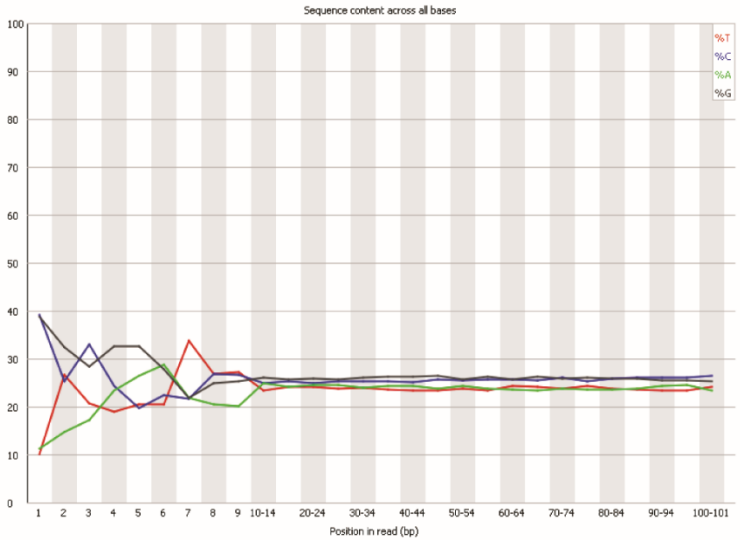
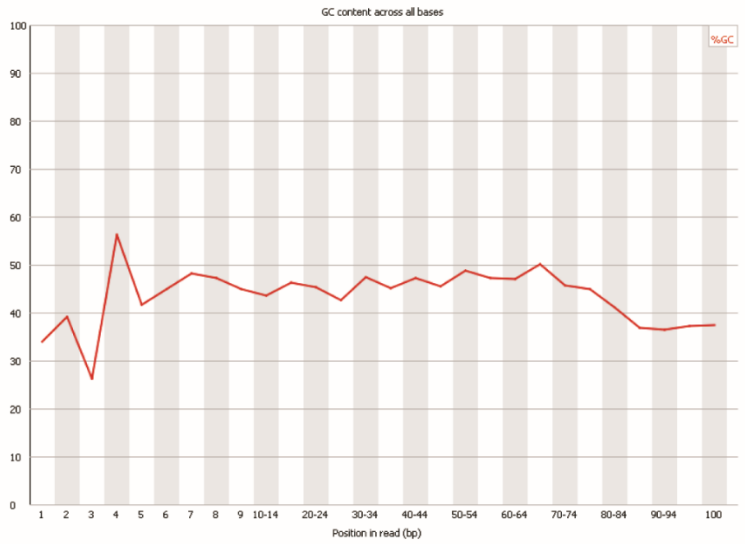


Figure 3.10 – Per base sequence content of SAF-A/hnRNP U CLIP sample and E14 input controls (libraries 1 and 2).

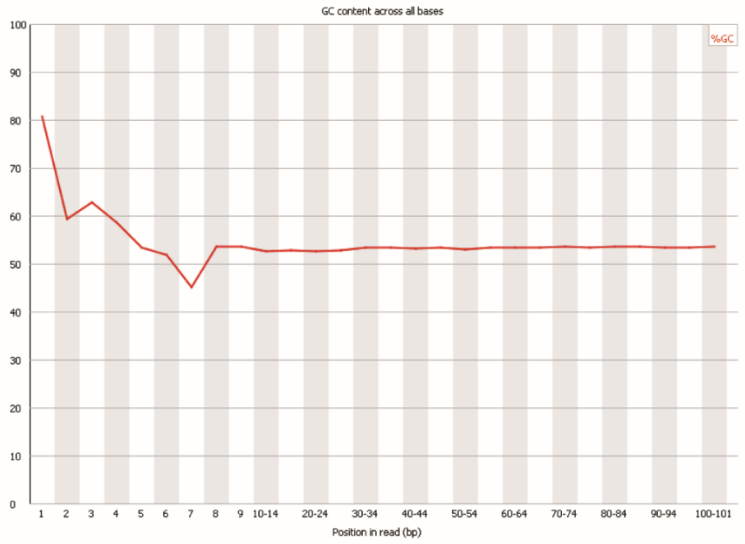
In the ideal case an even distribution of the four base positions should be seen – all four line should be parallel to one another. The strong variability in the beginning of each of the graphs can be explained by a sequence bias in the sequencing primers. Whereas the control libraries are rather unbiased the sample library shows slight biases for certain base pairs at some of the sequence positions. Furthermore, there is a remarkable increase of A towards the end of the SAF-A/hnRNP U CLIP sample.

After looking at the per base sequence content the per base GC content was examined (Figure 3.11). Also in this criterion the actual sample varies from the controls.

SAF-A/hnRNP U library



Control library 1



Control library 2

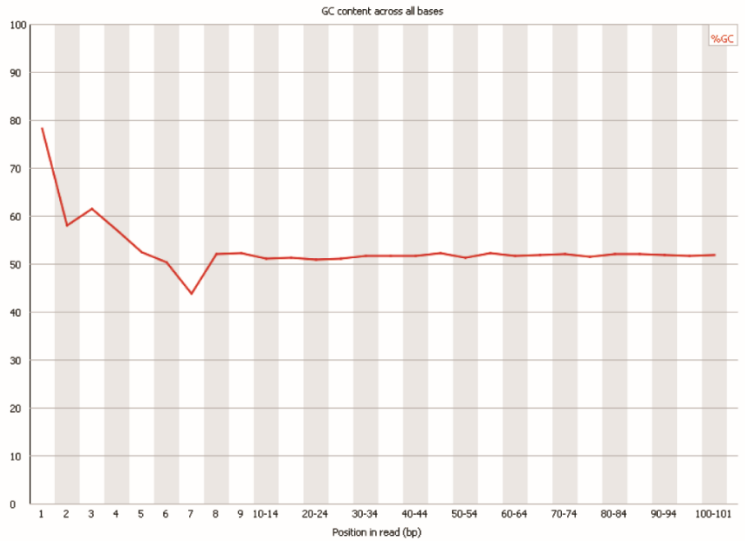
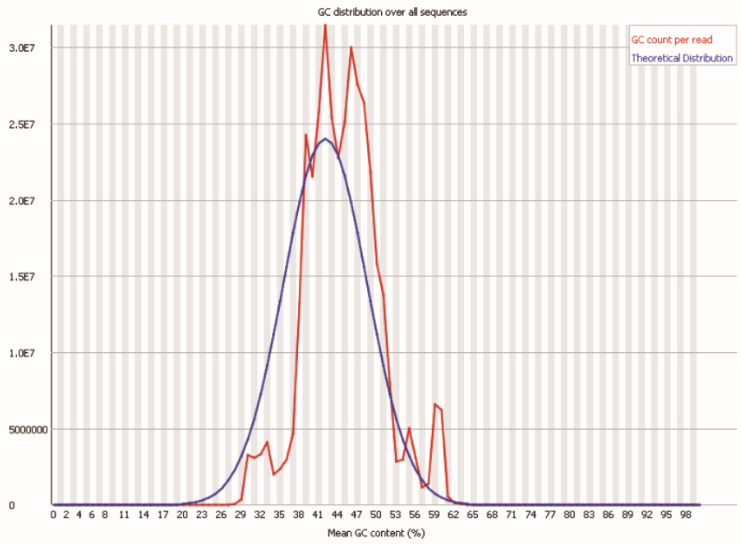


Figure 3.11 – Per base GC content in the SAF-A/hnRNP U CLIP library and E14 input controls (libraries 1 and 2).

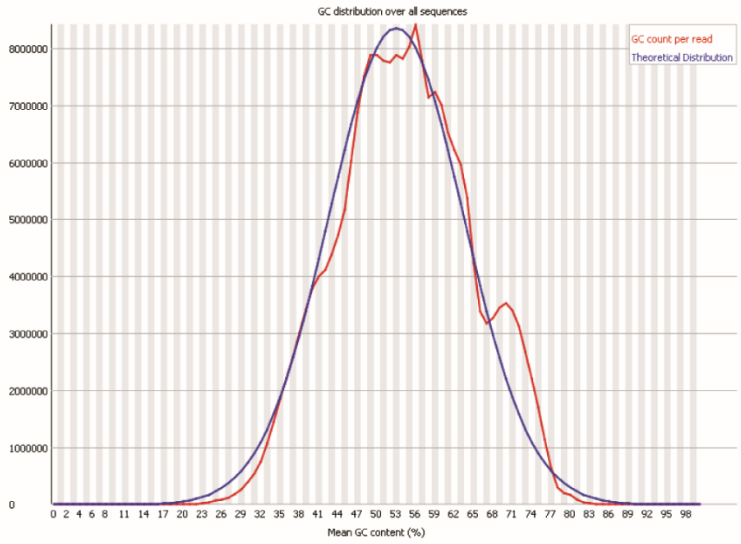
This plot is done to see whether the GC content changes across the reads in the different libraries. Once again the irregularities in the first 8 base reads are normal sequencing artifacts. Afterwards the graphs are usually expected to be fairly flat. This is different for the CLIP library where some variation is being observed.

Subsequently, the per sequence GC content was examined (Figure 3.12). In this analysis the GC content was measured along the entire sequence and compared to a modelled normal distribution (Andrews). A skew in the distribution may be an indicator for adapter contaminations, overrepresented sequences or for a contamination of the library with RNA from species other than mouse. Figure 3.12 shows that none of the libraries assumes a near-perfect normal distribution and once again the sample library with its sharp peaks is the furthest away from the normal distribution indicated by the blue line.

SAF-A/hnRNP U library



Control library 1



Control library 2

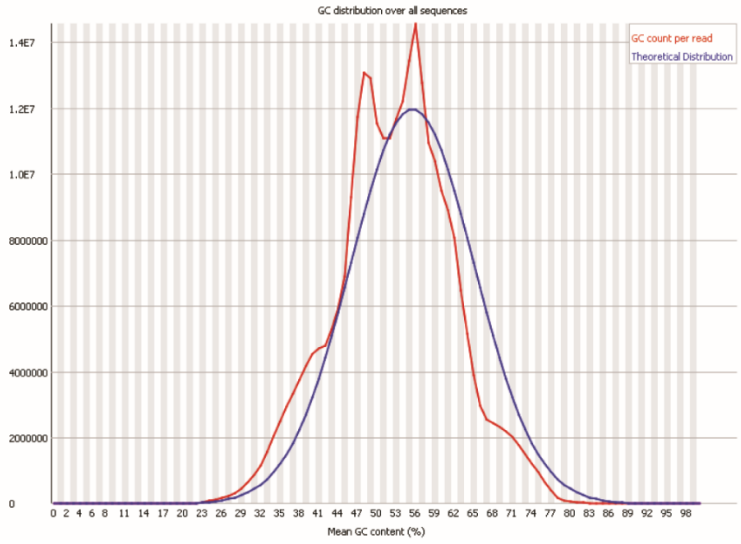
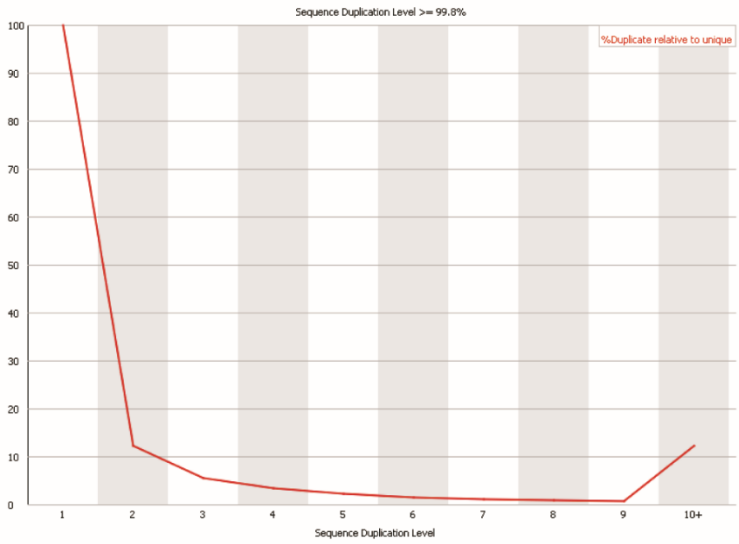


Figure 3.12 – Per sequence GC content of SAF-A/hnRNP U CLIP library and E14 input controls (libraries 1 and 2).

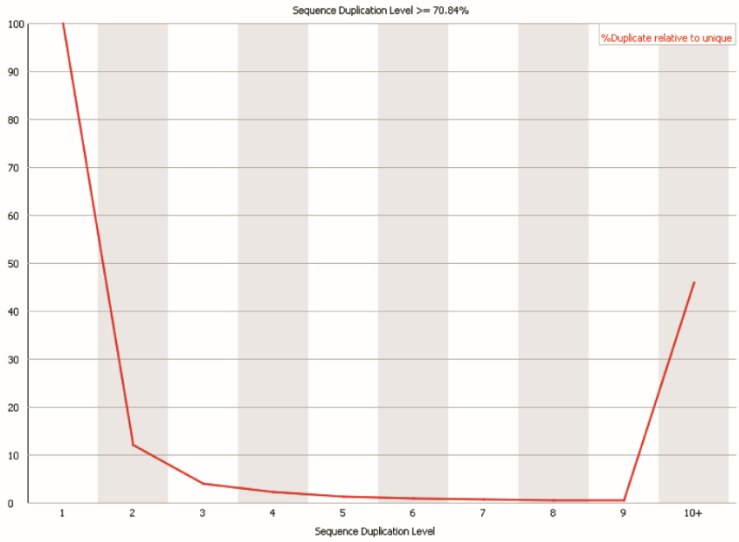
These are plots of the distributions of the GC content of all sequences where an ideal sample (red line) would assume a complete normal distribution (blue line). In the actual samples, however, there are secondary peaks coming out of the Gaussian curve indicating for some abnormalities within the libraries. The SAF-A/hnRNP U CLIP library shows the worst fit.

Importantly, sequence duplication levels were measured as well, and are depicted in Figure 3.13. Low levels of duplication usually indicate high levels of coverage, and high levels of duplication often indicate for an enrichment bias (e.g. PCR over-amplification) (Andrews). However, the underlying assumption of this model is that the library in question is unenriched. This is partially true for the control libraries as they were prepared from the total RNA of a set of E14 cell. Nevertheless, they have duplication levels of about 70%. These elevated levels are due to the fact that in RNA-Seq libraries different copy numbers of transcripts are present in the starting population (Andrews). To then observe lowly expressed transcripts researchers tend to greatly over-sequence highly expressed transcripts which leads to the generation of large amounts of duplicated sequences. The SAF-A/hnRNP U CLIP sample library, on the contrary, has a duplication level of 99.8%. When looking back at the TOPO cloning results discussed in Section 3.3 of this Chapter such high rate of duplication was not unexpected. Given the nature of the CLIP experiment it is difficult to tell whether this is a bias introduced through PCR amplification or a genuine enrichment of very specific sequences isolated by the SAF-A/hnRNP U procedure. Notably, all three libraries were not trimmed prior to analysis. Therefore, adapter contaminations inevitably contribute to the high duplication rates.

SAF-A/hnRNP U library



Control library 1



Control library 2

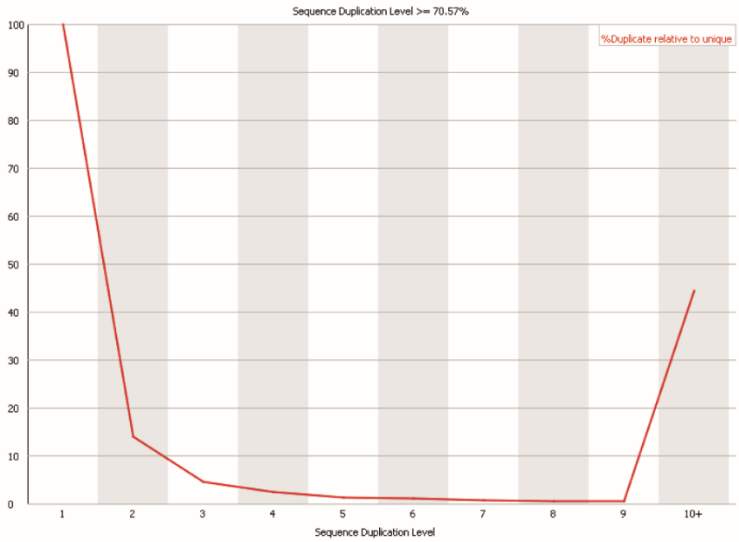


Figure 3.13 – Sequence duplication levels in the SAF-A/hnRNP U CLIP library and E14 input controls (libraries 1 and 2).

Sequence duplication levels are indicators of how unique sequences within a library are. Ideally a given sequence occurs in the library only once. As seen in the plots this is not the case for the RNA libraries presented here. The plot also provides information on sequence duplication levels that are high in all samples and close to 100% in the SAF-A/hnRNP U CLIP library.

In an attempt to understand the origin of the overrepresented sequences in SAF-A/hnRNP U CLIP library, a list of individual sequences that were present at especially high levels was prepared (Table 3.2). A brief glance over Table 3.2 shows that most of the duplications stem from the library preparation reagents (adapters, primers). In order to obtain the true duplication rate, these would need to be trimmed off the present library. But trimming is not a requirement for faithful alignment of the library to the mouse genome, as usually adapter/primer sequences cannot be aligned.

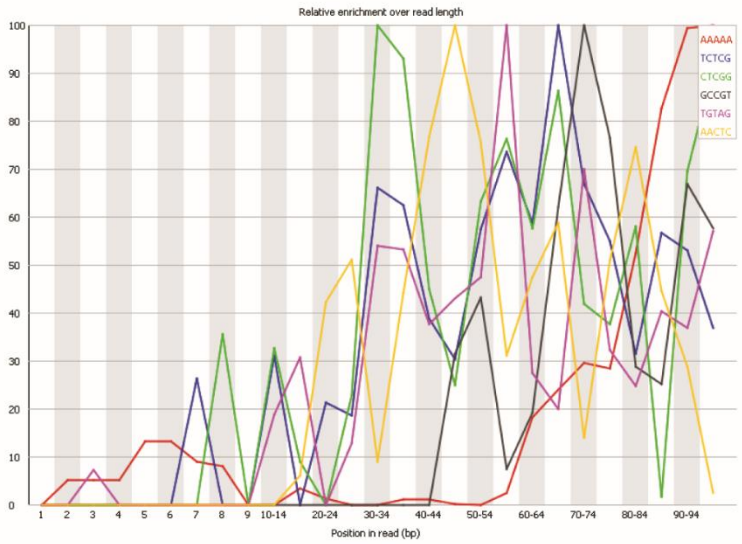
Table 3.2 – List of overrepresented sequences in the SAF-A/hnRNP U CLIP library.

Many of the overrepresented sequences in the library stem from library preparation reagents.

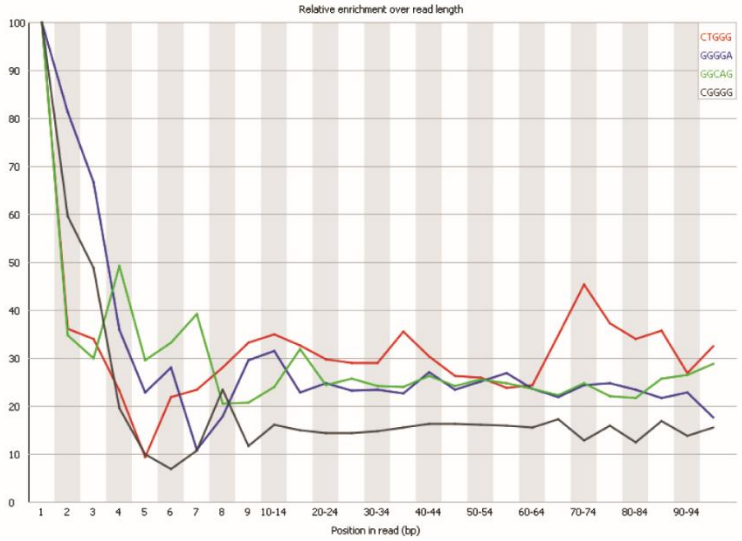
Sequence	Count	Percentage	Possible source
CCACGATCGTCGGACTGTAGAACTCTGAACGTGTAGATCTCGGTGGTCGC	7732489	1.982624955	Illumina RNA PCR Primer (100% over 41bp)
GTGGTGGAATTCTCGGGTGCCAAGGAACTCCAGTCACGGCTACATCTCGT	7219894	1.851194618	RNA PCR Primer, Index 11 (100% over 46bp)
ATACCACCCTGAACGCGCCCGATCTCGTCTGATCTCGGAAGCTAAGCAGG	6253280	1.603352941	No Hit
GCTCAAAAGAAAAGGCAAGACTTCTTACTATCTAATTGGTTATGACTATA	6213889	1.593253013	No Hit
ATAAGCTCCCGCTGTAGCCTCCGGAGTCCTGGAATTCTCGGGTGCCAAGG	6110415	1.566722082	Illumina Small RNA Adapter 2 (100% over 21bp)
TTTCAAAAAGCTATAGTCATAACCAATTAGATAGTAAGAAGTCTTGCCTT	6103919	1.565056495	No Hit
GGACTCCGGAGGCTACAGCGGGAGCTTATGATCGTCGGACTGTAGAACTC	6028014	1.545594308	Illumina DpnII expression Sequencing Primer (100% over 21bp)
AAATGCTGGAACGTGGTGTCAGTTCTGAACTCGGAAATGGAATTCTCGGG	5901386	1.513126647	No Hit
GGCGGTCTCCCATCCAAGTACTAACCAGGCCCGACCCTGCTTAGCTTCCG	5852246	1.500527057	No Hit
TTTCCGAGTTCAGAACTGACACCACGTTCCAGCATTTGATCGTCGGACTG	5769494	1.479309286	No Hit
CTGCAGCCACTTTGATGATCCCCTAACATCTGTAGCCATCATTCATCATC	5277758	1.353227236	No Hit
CCGAGTTCCAAATCCAAAGTCCAAAAATCTATATTCTTTCAAATAGATCG	5173938	1.32660759	No Hit
TATTTGAAAGAATATAGATTTTTGGACTTTGGATTTGGAACTCGGTGGAA	5171544	1.325993764	No Hit
ACTCCCAGCATGCAGGAGGCAGAGGCAGACGGATCTCTGAGTTGGAGGCC	4942868	1.2673608	No Hit
TAGCTAGTGAAAGAACATTCTGTTTCTTTCCTTTACCAGTAAGAGTACAT	4907444	1.258278019	No Hit
AGTTCTACAGTCCGACGATCAGATGATGAATGATGGCTACAGATGTTAGG	4800663	1.230899167	No Hit
TGTGGGGCTTAAAGGAAAATGAGGGTAAAGGGGAGCTAAGGAAAAATGAT	4737775	1.214774563	No Hit
ACTTTCTTGCTTGGCAGGCAAGAAGCCATATGGAATTCTCGGGTGCCAAG	4709714	1.207579669	No Hit
AAACATCGTACATACAATAAATAAATGGAATTCTCGGGTGCCAAGGAACT	4700631	1.205250771	RNA PCR Primer, Index 1 (100% over 25bp)
CCTCACACCAGTCAGAATGTCTAAGATCGTCGGACTGTAGAACTCTGAAC	4668900	1.197114882	Illumina DpnII expression Sequencing Primer (100% over 26bp)
TTTATTTATTGTATGTACGATGTTTGATCGTCGGACTGTAGAACTCTGAA	4621867	1.185055531	Illumina DpnII expression Sequencing Primer (100% over 25bp)
TATGGCTTCTTGCCTGCCAAGCAAGAAAGTGATCGTCGGACTGTAGAACT	4577685	1.173727181	No Hit
TACTTCTTTCTTTCTTTTTTTTTCCTTTTTTCTTTCTTTTCATGAGACAG	4526008	1.1604771	No Hit
GAACAGCCAGGGCTATACAGAGAAGATCGTCGGACTGTAGAACTCTGAAC	4520102	1.15896279	Illumina DpnII expression Sequencing Primer (100% over 26bp)
TTAGACATTCTGACTGGTGTGAGGTGGAATTCTCGGGTGCCAAGGAACTC	4423705	1.134246415	RNA PCR Primer, Index 1 (100% over 26bp)
TTCTCTGTATAGCCCTGGCTGTTCTGGAATTCTCGGGTGCCAAGGAACTC	4384948	1.124309046	RNA PCR Primer, Index 1 (100% over 26bp)
AAGAGAACAGTACAGTATTGGAATTCTCGGGTGCCAAGGAACTCCAGTCA	4186095	1.073322757	RNA PCR Primer, Index 1 (100% over 32bp)
ATACTGTACTGTTCTCTTGATCGTCGGACTGTAGAACTCTGAACGTGTAG	4152959	1.064826623	Illumina DpnII expression Sequencing Primer (96% over 30bp)
TAGACCAGAACCCTGCTCCTCCTGGAATTCTCGGGTGCCAAGGAACTCCA	4067583	1.042936054	RNA PCR Primer, Index 1 (100% over 28bp)
GGAGGAGCAGGGTTCTGGTCTAGATCGTCGGACTGTAGAACTCTGAACGT	3916840	1.004285261	Illumina DpnII expression Sequencing Primer (96% over 28bp)
CTGTATGCCATGGAGTGTGTCCACAAAGATCGTCGGACTGTAGAACTCTG	3846059	0.986136877	Illumina DpnII expression Sequencing Primer (100% over 23bp)
TTTGTGGACACACTCCATGGCATACAGTGGAATTCTCGGGTGCCAAGGAA	3751783	0.96196433	RNA PCR Primer, Index 1 (100% over 23bp)
TCTGTGCAGTGTTCACAGAGGCCTGATCGTCGGACTGTAGAACTCTGAAC	3670901	0.941226031	Illumina DpnII expression Sequencing Primer (100% over 26bp)
AGGCCTCTGTGAACACTGCACAGATGGAATTCTCGGGTGCCAAGGAACTC	3639772	0.933244496	RNA PCR Primer, Index 1 (100% over 26bp)
TCTCTTGTTCTTTTCTCTTCCCGTCAATGATCGTCGGACTGTAGAACTCT	3634570	0.931910693	Illumina DpnII expression Sequencing Primer (100% over 22bp)
GCATCAGTGAATATACACTCCCCTTTCCCACATCTATCTAGCATCAGTGA	3575354	0.916727598	No Hit
ATGTTAGAACCTGAAACTGGAATTCTCGGGTGCCAAGGAACTCCAGTCAC	3469383	0.889556431	RNA PCR Primer, Index 1 (100% over 33bp)
ATTGACGGGAAGAGAAAAGAACAAGAGATGGAATTCTCGGGTGCCAAGGA	3441891	0.882507429	RNA PCR Primer, Index 1 (100% over 22bp)
GTTTCAGGTTCTAACATGATCGTCGGACTGTAGAACTCTGAACGTGTAGA	3405754	0.873241834	Illumina DpnII expression Sequencing Primer (96% over 30bp)
ACTTTTGGAATTCTCGGGTGCCAAGGAACTCCAGTCACGGCTACATCTCG	3091426	0.792647534	RNA PCR Primer, Index 11 (100% over 45bp)
CCACGGATGGAATTCTCGGGTGCCAAGGAACTCCAGTCACGGCTACATCT	2904863	0.744812424	RNA PCR Primer, Index 11 (100% over 43bp)
AAAGTGATCGTCGGACTGTAGAACTCTGAACGTGTAGATCTCGGTGGTCG	2853946	0.731757208	Illumina RNA PCR Primer (100% over 40bp)
TTTACACATCTGCCAGCTCCATTAGACAAAGATCGTCGGACTGTAGAACT	2848654	0.730400329	No Hit
ACAAAGGATATGCTAAGGATAGGCTTAGCATCATGGAATTCTCGGGTGCC	2806221	0.719520427	No Hit
TCCGTGGGATCGTCGGACTGTAGAACTCTGAACGTGTAGATCTCGGTGGT	2763563	0.708582834	Illumina RNA PCR Primer (100% over 38bp)
TGATGCTAAGCCTATCCTTAGCATATCCTTTGTGATCGTCGGACTGTAGA	2744701	0.703746581	No Hit
TTTGTCTAATGGAGCTGGCAGATGTGTAAATGGAATTCTCGGGTGCCAAG	2699360	0.692121062	No Hit
CATTGTTGTTGACTGCTAGTACTGTGTATCTCACAAATGGAATTCTCGGG	2627081	0.673588588	No Hit
TGGAATTCTCGGGTGCCAAGGAACTCCAGTCACGGCTACATCTCGTATGC	2331961	0.597919256	RNA PCR Primer, Index 11 (100% over 50bp)
GATCGTCGGACTGTAGAACTCTGAACGTGTAGATCTCGGTGGTCGCCGTA	2243619	0.575268199	Illumina RNA PCR Primer (100% over 45bp)
AATTGATGACTTAGATGTGGGGTAAAAGGTAAATGGAATTCTCGGGTGCC	933427	0.239332466	No Hit
TTTACCTTTTACCCCACATCTAAGTCATCAATTGATCGTCGGACTGTAGA	932816	0.239175805	No Hit
AGATGATGAATGATGGCTACAGATGTTAGGGGATCATCAAAGTGGCTGCA	509266	0.130576776	No Hit
TTGGTGGAATTCTCGGGTGCCAAGGAACTCCAGTCACGGCTACATCTCGT	415253	0.106471663	RNA PCR Primer, Index 11 (100% over 46bp)

Lastly, Kmer profiles for the three libraries in question were prepared (Figure 3.14). Notably, only a proportion of the library (2%) is used and from these data the Kmer content is extrapolated to the remainder of the library (Andrews). Figure 3.14 shows that the Kmer profiles of both control libraries look relatively similar. On the other hand the SAF-A/hnRNP U library Kmer graph looks very different. As mentioned previously these differences may be due to variations in library preparation or experimental setups.

SAF-A/hnRNP U library



Control library 1



Control library 2

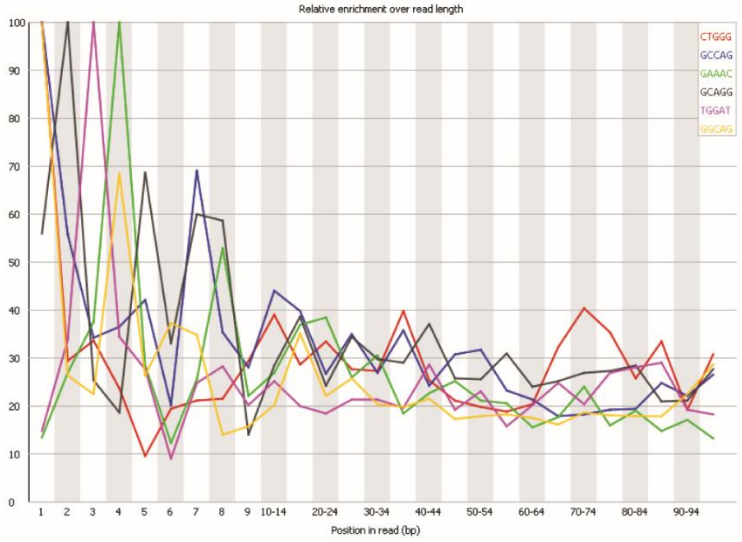


Figure 3.14 – Kmer profiles for SAF-A/hnRNP U CLIP library and E14 input controls (libraries 1 and 2).

Here the enrichment of every 5-mer within a sequence has been counted. The graphs show the enrichments of the top Kmers along the length of the read. In the case of CLIP Kmers could provide information on the SAF-A/hnRNP U binding consensus sequence.

Given the experimental constraints entailed by a CLIP experiments the quality of the SAF-A/hnRNP U CLIP dataset as assessed by FastQC appears sufficient for subsequent analysis. Especially the sequences of enriched Kmers that do not correspond to Illumina primer or adapter sets will be of interest in futures analyses as they can provide information on the SAF-A/hnRNP U RNA binding consensus sequence.

When looking back at the individual quality control steps it is important to bear in mind that *de facto* the E14 input control and the SAF-A/hnRNP U CLIP libraries are not entirely comparable because: 1) The libraries were prepared from different ESC lines. 2) The input library comprises the total cellular RNA. In contrast, the CLIP library represents a restricted set of RNAs by default. 3) The libraries were not prepared in parallel.

A potential problem of the CLIP library is the high level of duplicates that may either be PCR artifact (technical duplicates) or represent a true biological situation (biological duplicates). On the sequence level there is no way to distinguish between the two. However, the low amount of input RNA, that is in fact lower than the concentrations commonly recommended by manufacturers of high-throughput sequencing library preparation kits, increases the chances that highly abundant sequences are over-amplified.

Since these open questions could not be addressed with the present datasets and the bioinformatics tools at hand, the preliminary results presented here should be interpreted with caution.

3.5 Qualitative SAF-A/hnRNP U CLIP-Seq data analysis

A qualitative, visual inspection of the SAF-A/hnRNP U CLIP dataset in the Integrated Genomics Viewer (IGV) revealed that the majority of reads located to intronic regions of genes or to intergenic/non-annotated loci. Figure 3.15D shows an exemplary intronic binding site of SAF-A/hnRNP U to Ptprm RNA. Unexpectedly, no reads were found to align with the Xist locus (Figure 3.15A). Instead there were both intron and exon binding sites on another lncRNA (A830018L16Rik/lnc-AGTRAP-1:1) with a yet unknown function (Figure 3.15B and C).

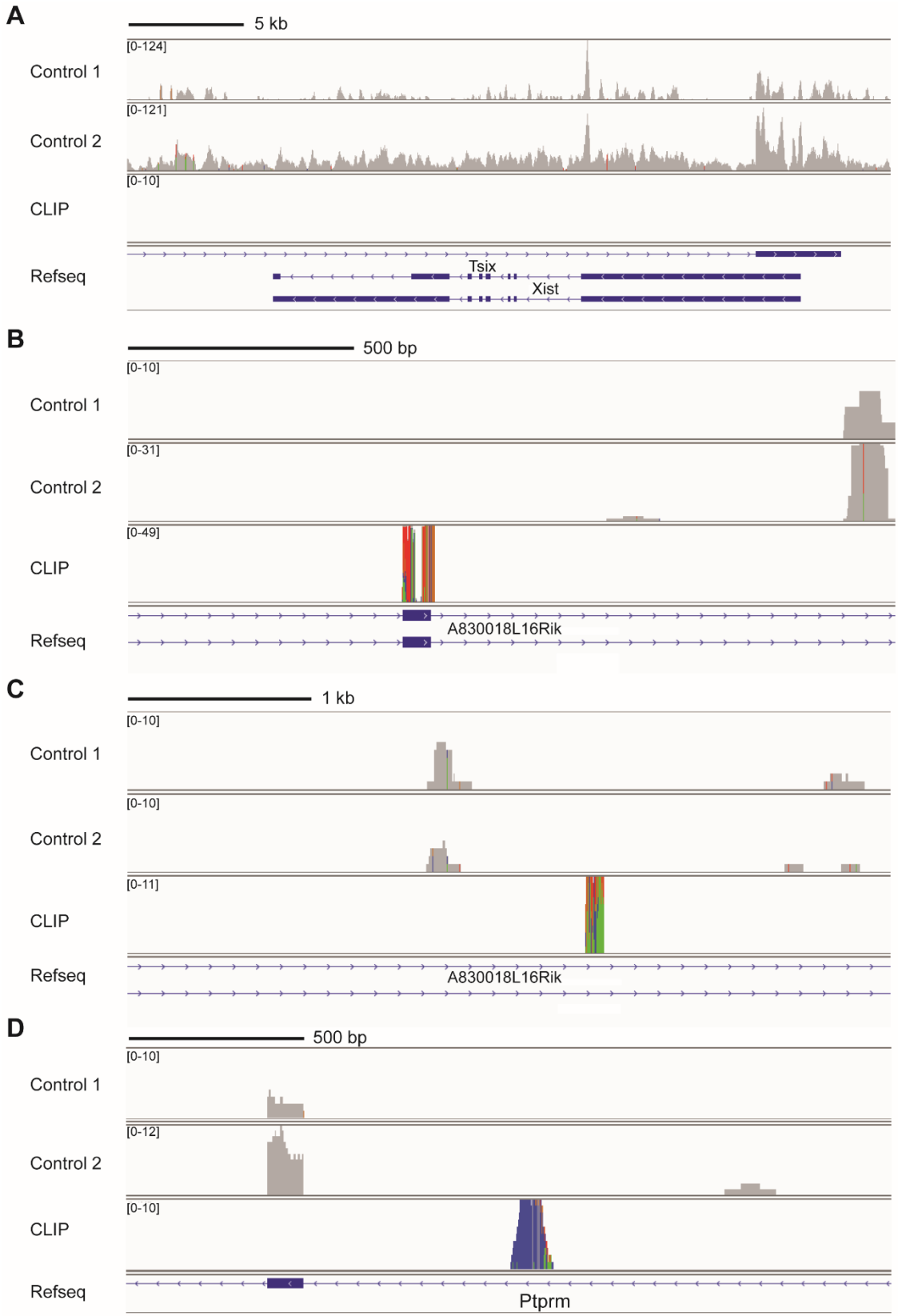


Figure 3.15 – IGV profiles of mouse ESC input controls and the SAF-A/hnRNP U sample.

Sequencing profiles of mouse input controls (control 1 and 2) and the SAF-A/hnRNP U CLIP (CLIP) at selected loci (Refseq genes) are shown. Blue boxes represent exons and blue arrowed lines represent intronic regions. The orientation of the arrowheads provides information on the direction of transcription. A scale bar is presented above each screenshot. (A) Xist/Tsix locus with reads in the control, but no reads in the SAF-A/hnRNP U CLIP sample. (B) SAF-A/hnRNP U binding to exon 10 on of the lncRNA A830018L16Rik (lnc-AGTRAP-1:1). (C) SAF-A/hnRNP U binding to intron 11 of the lncRNA A830018L16Rik (lnc-AGTRAP-1:1). (D) Peak in an intronic region (intron 9) of the Protein tyrosine phosphatase receptor type M (Ptprm).

The SAF-A/hnRNP U intron, intergenic region and lncRNA associations are in agreement with the literature. In 2012, Xiao *et al.* published HeLa cell SAF-A/hnRNP U CLIP-Seq data showing that in this particular model system 57% of the CLIP tags mapped to introns, 13.4% to intergenic regions and 12% to annotated non-coding RNAs (Xiao, Tang et al. 2012). However, due to inter-species differences between human and mouse no direct comparisons can be made between the Xiao *et al.* study and the present work. In Hasegawa's report the female mouse neuroblastoma cell line Neuro2a had been utilized. As the main focus of our work was on ESCs, a male mouse ESC lin bearing an inducible Xist transgene had been utilized. This cell line had been selected because it opened up the opportunity to directly compare the Xist-induced and uninduced state in the same cellular system. Thereby, strong upregulation of Xist RNA may have eased the dectection of direct RNA-protein interactions.

This is different in the Hasegawa *et al.* study, where an association of SAF-A/hnRNP U with Xist RNA had been shown in the mouse model system (Hasegawa, Brockdorff et al. 2010). The present study failed to confirm these findings, and there are many possible explanations for this observation. Firstly, different experimental approaches were taken in these two studies. Hasegawa *et al.* used a crude RNA IP method followed by qRT-PCR, whereas the present study utilized a refined technique that specifically purifies SAF-A/hnRNP U-RNA

complexes. Therefore, there is the possibility that the previously observed SAF-A/hnRNP U-Xist RNA interactions were an experimental artifact. However, it may also be that the interactions between SAF-A/hnRNP U and Xist RNA are very transient or that so few interactions occur that the stringent CLIP technique cannot capture these. Another difference between the two approaches, that may be the origin of the divergent results, is the usage of two different cell lines. In Hasegawa's report the female mouse neuroblastoma cell line Neuro2a had been utilized. For this thesis work, on the contrary, a male mouse ESC carrying an inducible Xist-transgene was utilized to prepare the input samples for the CLIP experiment. Variations between a differentiated model system, where XCI had already been established (Neuro2a), and a mouse ESC model, where Xist expression can be induced from a transgenic locus (3E), may be the reason for deviations in experimental outcomes. Furthermore, it should be reiterated that the CLIP-Seq results presented here are only preliminary, and more experiments are needed to allow an accurate interpretation of the extent to which SAF-A/hnRNP U can directly associate with Xist RNA. In addition, to gain a better understanding of the underlying Biology, the bioinformatic analysis for this and future CLIP datasets needs to be streamlined and improved on.

3.6 Summary and discussion

3.6.1 CLIP-Seq experimental considerations

Here an approach to optimizing SAF-A/hnRNP U CLIP conditions was shown. As CLIP-Seq experiments are very sensitive to varying experimental conditions, it is essential to always return to testing the individual steps of the protocol after changing an essential parameter. Furthermore, like every other technique, CLIP comes with certain general constraints, i.e. the efficiency of the pull-down strongly depends on structural parameters such as the positioning of reactive amino acids and bases and secondary structures. Therefore, an unfavorable protein and/or RNA configuration often results in low efficiencies of crosslinking (Sei and Conrad 2014). But even if a site was readily crosslinked, the experiment may not be

successful because the covalent linkage between RNA and protein may complicate the down-stream analysis. After completion of the protein-RNA complex purification, the proteinaceous component is digested away. However, oftentimes the proteolytic digest is incomplete leaving protein adducts attached to the RNA. These adducts impede reverse transcription by stalling the reverse transcriptase (Sei and Conrad 2014). Hence, it is inevitable that the cDNA amplification step, which is part of every CLIP protocol, introduces a bias into the sequencing output. The effect of this is aggravated by the low amounts of input RNA available for the reverse transcriptase reaction, and inefficient RNA adapter ligation reactions that often limit cDNA library complexity (Konig, Zarnack et al. 2011).
Ways to circumvent at least some of the problems entailed by conventional CLIP experiments have been developed a few years ago. Individual-nucleotide resolution CLIP (iCLIP) is one approach where the RNA ligation had been substituted by a more efficient cDNA circularization step (for more details see Konig, Zarnack et al. 2010), and it also allows to identify the exact positioning of the crosslinking site at the level of nucleotides (Konig, Zarnack et al. 2011). Another, improved, alternative to conventional CLIP is Photoactivatable Ribonucleoside-enhanced CLIP (PAR-CLIP). With the development of PAR-CLIP the efficiency issue of conventional CLIP had been addressed (Spitzer, Hafner et al. 2014). By incorporating photoreactive nucleotide analogues like 4-thiouridine (4SU) and 6-thioguanosine into the cellular RNA, proteins can be crosslinked to RNA with 365 nm UV light. The analogues enhance crosslinking efficiency in this setup. Furthermore, they allow for an easier identification of crosslinking sites because these are marked by characteristic mutations induced by the reverse transcription of nucleotide analogue-containing RNA.

In light of the most recent developments in the CLIP field, it would be worthwhile to also try SAF-A/hnRNP U iCLIP or photoactivatable ribonucleoside-enhanced CLIP PAR-CLIP. iCLIP overcomes two of the disadvantages associated with conventional CLIP where the identification of binding sites relies on the analysis of overlapping sequence clusters (Konig,

Zarnack et al. 2010). Therefore, distances less than 30 nucleotides cannot be resolved. Furthermore, CLIP requires the reverse transcriptase to transcribe over residual amino acids that remain attached to the RNA at the crosslinking site. Primer extension essays have shown that the transcriptase often stalls at these sites such that truncated cDNAs are generated (Urlaub, Hartmuth et al. 2002). iCLIP utilizes this apparent limitation by capturing truncated cDNA products utilizing a self-circularization process. Moreover, the introduction of a random barcode adapter in iCLIP allows for the discrimination between unique cDNAs and PCR duplicates which improved the qualitative information retrieved from the experiment.

PAR-CLIP on the other hand utilizes the nucleotide analogue 4SU (Hafner, Landthaler et al. 2010). Cells are grown on a 4SU-enriched medium such that an enrichment of the nucleotide analogue within the host RNA is achieved. Experimental evidence shows that by using this method a better enrichment and thus higher yield of crosslinked RNA is achieved. Additionally, the crosslinking site can be located by identifying T to C transitions in the cDNA of isolated RNA elements.

Irrespective of that, additional controls need to be included in all future experiments. If conventional CLIP was utilized, a knock-down control would be ideal. However, all knock-down trials conducted for this thesis work were unsuccessful (data not shown). In addition to that input controls of exactly the same ESC line should be sequenced in the next run, and even more importantly 3E cells without doxycycline treatment need to be analyzed as well. Furthermore, it should be noted that there is the possibility that, due to its structure and base composition, Xist RNA cannot be extracted by CLIP. Consequently, the progress in the field may entirely depend on the advancement of experimental techniques.

There is another possible optimization step that has not been taken into consideration here: The endogenous SAF-A/hnRNP U antibody does not highlight the protein's Xi enrichment in IF microscopy. One of the possible explanation for this observation may be epitope masking.

The masking may have been maintained in the CLIP input samples. In order to circumvent this caveat chromatin extraction procedures under denaturing conditions, i.e. 8M urea, should be tried in another set of experiments (Balamotis, Tamberg et al. 2012). The harsh extraction conditions may be able to exposed hidden epitopes and facilitate the recovery of SAF-A/hnRNP U-Xist complexes.

Apart from these purely technical parameters the choice of the biological model system should be rethought as well. In the current set of experiments male, Xist-inducible ESCs (3E) have been used. The system used is very artificial and may not give any relevant results due to the non-physiological nature of Xist expression and localization in this cell line. Therefore, all future experiments should also include a female cell line where abnormalities in Xist expression can be excluded.

3.6.2 SAF-A/hnRNP U RNA binding behavior

The preliminary CLIP experiment using Xist-induced 3E ESCs suggests that SAF-A/hnRNP U mainly binds to intronic and intergenic/non-annotated RNAs. As mentioned before this finding is in line with the data derived from another CLIP experiment (Xiao, Tang et al. 2012). Unlike here, Xiao *et al.* investigated the binding behavior of human SAF-A/hnRNP U in a cancer cell model (HeLa) without an Xi because the focus was on the involvement of the protein in splicing control. The apparent similarities between the two studies imply that there is a more general role for SAF-A/hnRNP U in the nuclei of mouse and human cells, i.e. for transcriptional regulation, splicing, structural organization and localization of nuclear RNAs.

This thesis work, however, was mainly concerned with linking SAF-A/hnRNP U to XCI, and it was hoped to confirm and strengthen Hasegawa *et al.*'s evidence showing that there was a direct interaction between SAF-A/hnRNP U and Xist RNA. In the CLIP experiment it could not be shown that the two macromolecules interact with one another. As mentioned in the results Section 3.5 of this Chapter, the outcomes of both experiments could have been

influenced by the conditions under which they were conducted. But there remains the possibility that the results presented here are genuine which would mean that SAF-A/hnRNP U is not a direct Xist-interacting protein.

At this point it is important to remember that other studies have shown profound differences between the nuclear matrices of cancer cells, undifferentiated and differentiated cells (Munkley, Copeland et al. 2011). Munkley *et al.* (2011) showed a differential association of cyclin E with the nuclear matrix in several cancer cell lines, and undifferentiated and differentiated cells. (Munkley, Copeland et al. 2011). Their finding lead the authors of this study to the conclusion that cancer cells are a questionable model system for studies of the nuclear matrix. Interestingly, Munkeley *et al.* also discovered that protein associations with the nuclear matrix may change in the course of differentiation. Thus, an altered matrix structure may be involved in the maintenance of the undifferentiated state (Munkley, Copeland et al. 2011).
In respect to the experimental design of the CLIP experiment presented in this thesis, this may imply that by default a SAF-A/hnRNP U association with Xist RNA cannot be seen, because the selected ESC line had remained undifferentiated even in the Xist-induced state.

Lastly, it should not be forgotten that CLIP is an antibody-dependent method. For the SAF-A/hnRNP U experiment the in house antibody against the endogenous protein characterized in Chapter 2 was used. Experiments in this Chapter also showed that the enrichment of SAF-A/hnRNP U at the Xi cannot be visualized using this reagent. The incapacity of this antibody to highlight the Xi may hint at that it does not bind to Xist-associated SAF-A/hnRNP U.

Chapter 4 describes the generation of a female somatic cell line stably expressing a green fluorescent protein (GFP)-tagged SAF-A/hnRNP U variety. In this cell line an enrichment of

SAF-A/hnRNP U was clearly visible. Therefore, future CLIP experiments should be done using this cell line.

4. SAF-A/hnRNP U stably associates with the Xi

4.1 Introduction

In 2003 Helbig and Fackelmayer published the first study that pointed towards a link between SAF-A/hnRNP U and the XCI process by showing that the N-terminally fluorophore-tagged protein was enriched at the human Xi (Helbig and Fackelmayer 2003). Unlike Fackelmayer's antibody-based approach, that necessitated an *in situ* nuclear matrix preparation step prior to detection, direct imaging of cells transfected with green fluorescent protein (GFP)-tagged SAF-A/hnRNP U also showed a clear enrichment of the protein at the Xi. Furthermore, SAF-A/hnRNP U enriched foci overlapped with the prominent Xi-marker histone variant macroH2A. Helbig and Fackelmayer's microscopy study built up on biochemical reports demonstrating that members of the hnRNP family can bind to XIST RNA *in vitro*, and that the RGG domain is essential for SAF-A/hnRNP U localization to the Xi (Brown and Baldry 1996). These initial studies using GFP-tagged SAF-A/hnRNP U were followed up on by others using the same tagging approach (Fackelmayer 2005, Pullirsch, Hartel et al. 2010). For instance, Pullirsch *et al.* (2010) complemented their autoimmune serum-based investigations with images of C-terminally GFP-tagged SAF-A/hnRNP U (Pullirsch, Hartel et al. 2010). Common to all studies was that the Xi-enrichment of fluorophore-tagged versions of the protein could readily be visualized in conventional fluorescence microscopy. Since the antibody generated and described in this thesis (see Chapter 2) failed to highlight the Xi-enrichment of SAF-A/hnRNP U at the Xi, GFP-tagged protein was studied as described in the following sections.

4.2 hrGFP-tagged SAF-A/hnRNP U is enriched at the Xi

For this purpose FSPE cell lines stably expressing humanized *Renilla reniformis* GFP (hrGFP) fused to the C-terminus of SAF-A/hnRNP U were generated (Figure 4.1). As published previously, hrGFP-tagged SAF-A/hnRNP U is enriched at the Xi that is marked by H3K27me3 (Figure 4.1A) (Pullirsch, Hartel et al. 2010). Anti-SAF-A/hnRNP U and anti-

hrGFP Western blotting showed that, even though the mouse cells appeared bright under the microscope, the fusion protein was underexpressed as compared to endogenous SAF-A/hnRNP U (Figure 4.1C). Densitometric analysis of the blot showed that that clone A7 (cA7) reached the highest expression with over 25% of endogenous levels. In the given experimental set up this is of advantage because problems such as cell death due to cytotoxicity and non-physiological protein behavior, which commonly arise from protein overexpression, are unlikely (Gibson, Seiler et al. 2013).

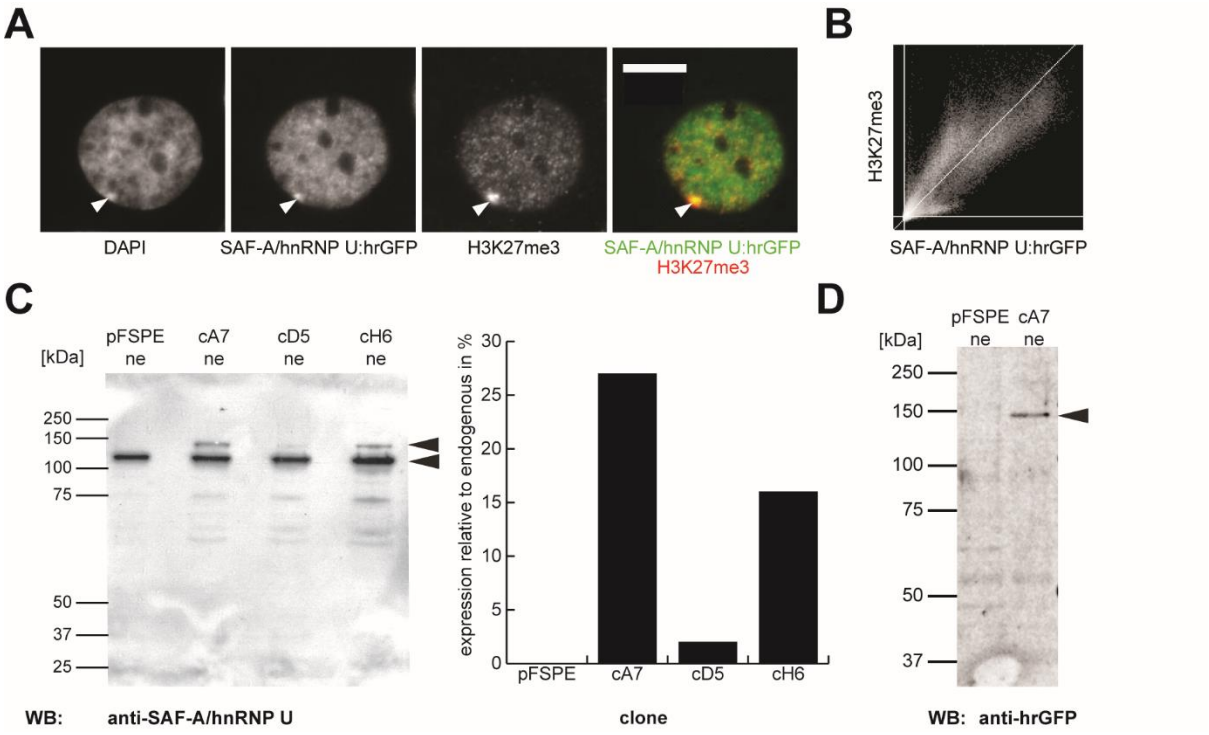


Figure 4.1 – SAF-A/hnRNP U:hrGFP stable cell line preparation.

(A) Representative IF images of a FSPE cell stably expressing SAF-A/hnRNP U:hrGFP. The white arrowheads (▼) point at the Xi seen in the 4', 6-diamidino-2-phenylindole (DAPI), the SAF-A/hnRNP U:hrGFP and the H3K27me3 channels. Scale bar 10 µm. (B) Intensity plot of SAF-A/hnRNP U:hrGFP against H3K27me3 shows a good correlation between the two. (C) Anti-SAF-A/hnRNP U Western blot on nuclear extracts (ne) from three selected clones (cA7, cD5 and cH6) and the parental cell line (pFSPE). Arrowheads (▼) point at the endogenous

protein and the hrGFP-tagged variant (lower and higher molecular bands, respectively). The bar chart shows expression levels of fluorophore-tagged protein in the various lines relative to endogenous expression as determined by densitometry on the Western blot shown on the right hand side. (D) Anti-hrGFP blot on nuclear extracts from pFSPEs and a selected clone (cA7) showing a single band for hrGFP-tagged SAF-A/hnRNP U (▼).

Subsequently, cells stably expressing the fluorescently tagged SAF-A/hnRNP U proteins were tested in Immuno-Xist RNA fluorescence *in situ* hybridization (FISH) experiments (Figure 4.2). These showed that SAF-A/hnRNP U:hrGFP and Xist RNA FISH signals overlap in conventional fluorescence imaging on fixed cells. However, there is a clear difference between the Xist FISH images in the green and the red microscopy channels (displayed in greyscale in Figure 4.2; left and right panels, respectively). When using the red-labelled probe Xi-enriched Xist RNA and the usual cytoplasmic background were observed. On the contrary, cells incubated with the green-labelled probe exhibited some background signal originating from the hrGFP-tagged protein. This finding was rather surprising as the denaturing agent formamide had been used at the Xist probe hybridization stage that preceded the immunolabelling step with hrGFP antibody (Nakamura, Kameda et al. 2008). This indicates that not all hrGFP molecules had been denatured resulting in an incomplete quenching of the green SAF-A/hnRNP U:hrGFP signal. Therefore all future experiments should be done with red-labelled Xist RNA probes, with SAF-A/hnRNP U being imaged in the green channel.

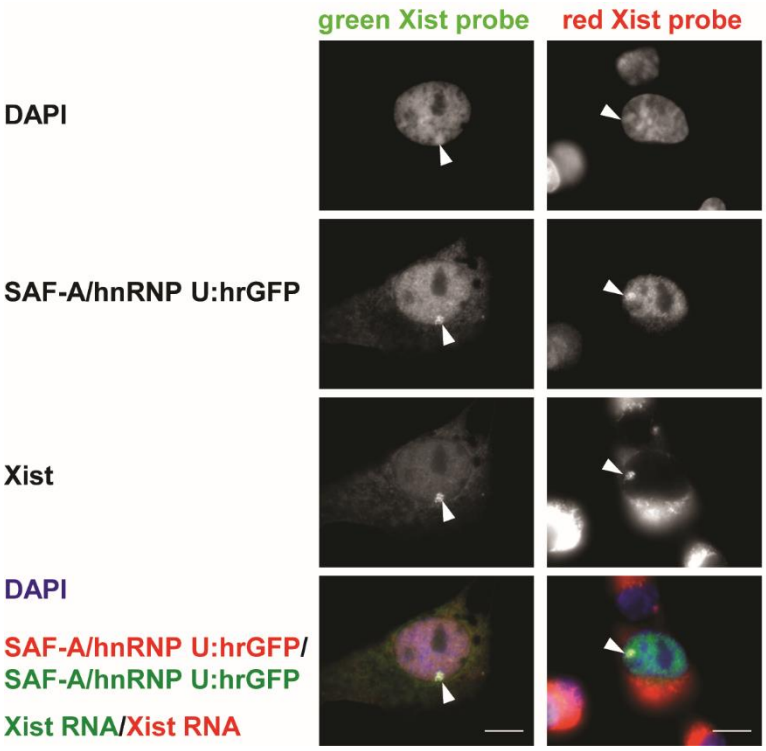


Figure 4.2 – Co-localization of Xist RNA and Xi-enriched hrGFP-tagged SAF-A/hnRNP U.

Xist RNA immuno-fluorescence *in situ* hybridization (FISH) on FSPE cells stably expressing SAF-A/hnRNP U:hrGFP. The protein was stained using a polyclonal hrGFP antibody. For specificity testing the color labels of probe and secondary antibody were swapped between the two experiments. Individual channels are shown in greyscale, merged images in color with DAPI in blue, SAF-A/hnRNP U:hrGFP in red (left hand panel) and green (right hand panel), Xist RNA in green (right hand panel) and red (left hand panel). Positions of the Xi in individual channels are indicated by arrowheads (▼). Scale bars 10 µm.

4.3 SAF-A/hnRNP U:hrGFP stably associates with the Xi

An earlier study reported that GFP-tagged, human SAF-A/hnRNP U stably associates with the interphase Xi (Fackelmayer 2005). The present work, however, focuses on SAF-A/hnRNP U in the mouse model. Because it could not be ruled out that there were inter-

species differences, the stability of murine SAF-A/hnRNP U:hrGFP-Xi associations in interphase nuclei were analyzed.

A variety of non-invasive, imaging-based methods are commonly used to investigate intracellular protein dynamics, i.e. fluorescence recovery after photobleaching (FRAP) and fluorescence correlation spectroscopy (FCS) (Sprague and McNally 2005). Whereas FRAP looks at events that occur within seconds, FCS is better suited to faster events that are on a scale of microseconds. Since previously SAF-A/hnRNP U dynamics were in the range of seconds, FRAP was the method of choice here (Fackelmayer 2005). Briefly, FRAP measures the rate of fluorescence recovery at a photo-bleached site. For this purpose fluorescent molecules in the region of interest are irreversibly damaged by a bleaching laser beam (c.f. Figure 4.6B) (Sprague and McNally 2005). The empirically determined fluorescence recovery rate then provides a measure for how quickly unbleached molecules move back into the bleached territory. To understand the outcome of a given FRAP experiment it is important to know that a protein's mobility is determined in part by diffusion of the molecules through the intracellular environment, and in part by binding interactions that detain the proteins from diffusing freely.

Here the regions of interest were the Xi-associated- and the nuclear pool of SAF-A/hnRNP U (Figure 4.3A). Both pools were spot bleached and the recovery of fluorescence was quantified using the image processing software Fiji (Figure 4.3B). It is apparent that the nuclear pool recovered nearly completely within the 600 seconds of imaging time. The Xi-pool, on the contrary, plateaus very early on, and recovered only up to about 30% at the end of the imaging time.

To complement this finding the reverse approach was taken. In an inverse FRAP (iFRAP) experiment the entire nuclear protein pool was quenched leaving out the Xi-associated SAF-

A/hnRNP U. The setup aimed at investigating the diffusion behavior of SAF-A/hnRNP U away from the Xi (Figure 4.3C). Also in this setup SAF-A/hnRNP U remained stably associated with the Xi and did not diffuse away into the nucleoplasm.

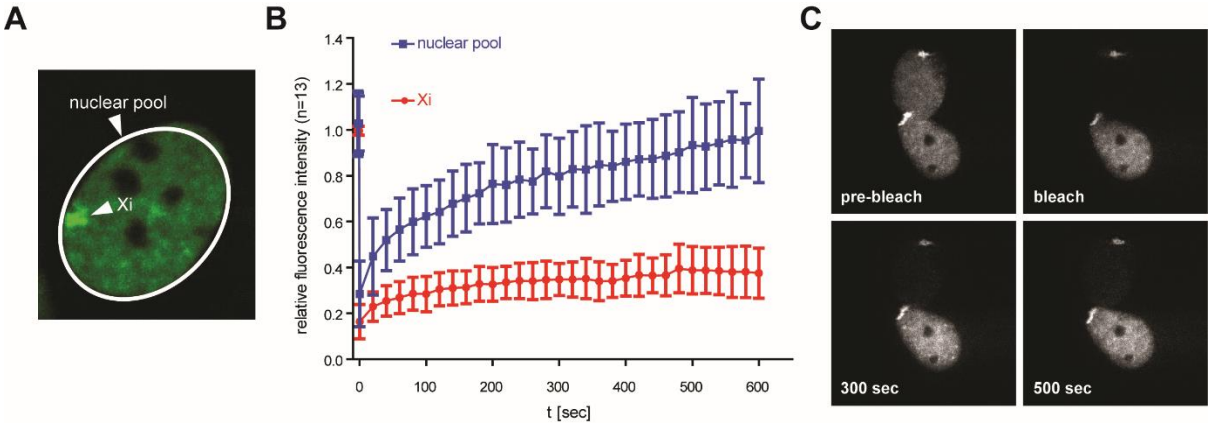


Figure 4.3 – Stable association of SAF-A/hnRNP U with the Xi.

(A) FSPE nucleus with hrGFP-tagged SAF-A/hnRNP U. Nuclear and Xi-associated pools are indicated by arrowheads (▼). (B) Fluorescence recovery after photobleaching (FRAP) curves for photobleaching experiments on the nuclear and the Xi-associated pools in blue and red, respectively. Curves show that over time the nuclear pool recovers almost completely whereas the Xi-associated pool does not (time in seconds [sec], number of cells n=13, error bars represent the standard deviation). (C) Inverse FRAP experiment (iFRAP) on the same cell line showing that SAF-A/hnRNP U remains stably associated with the Xi. An unbleached cell is shown for comparison (bottom cell). Scale bar 10 µm.

Figure 4.4 shows individual time-lapse images of a single cell that is a representative of the 13 cells making up the FRAP curve presented in Figure 4.2B. Another unbleached cell was followed in the same field of view to control for adverse effects due to laser irradiation. The Xi of the control cell is not visible because it locates to a different plane. The image series shows that the laser intensity used in this study does not cause any obvious damage like changes in nuclear morphology or cell motility.

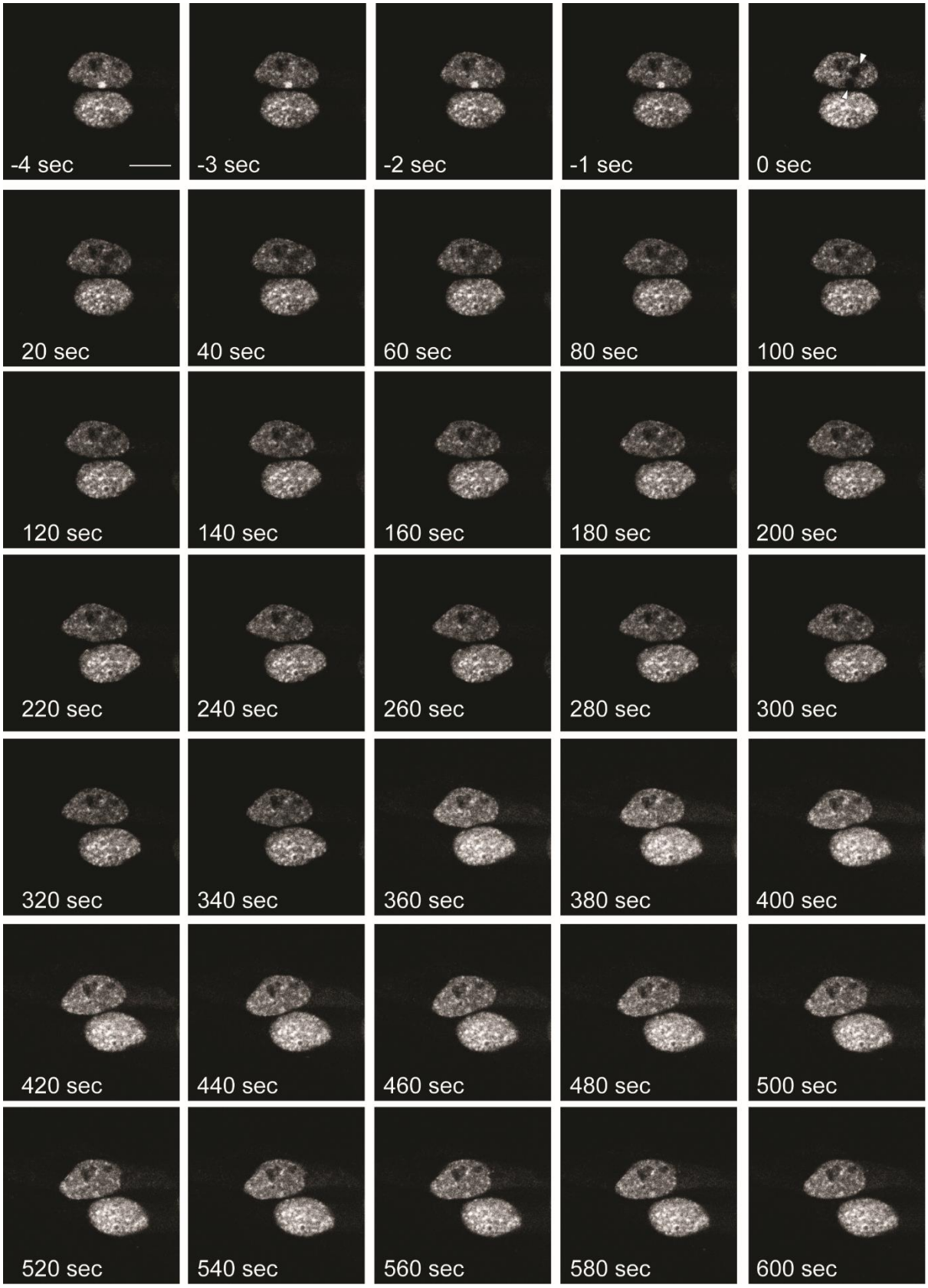


Figure 4.4 – Exemplary SAF-A/hnRNP U FRAP image series.

Images taken before (-4 to -1 sec) during (0 sec) and after photobleaching (20 to 600 sec) of SAF-A/hnRNP U:hrGFP in FSPE cells. For control purposes two cells were imaged at a time. The cell with the SAF-A/hnRNP U enrichment at the Xi in focus (upper cell) was subjected to photobleaching. The Xi of the other cell (lower cell) was located in another focal plane. The spot-bleached sites are indicated by arrowheads (▼) at 0 sec. All subsequent images were taken at 20 second intervals. Scale bar 10 µm.

4.4 Summary and discussion

4.4.1 SAF-A/hnRNP U:hrGFP localization in FSPE cells

In line with other published works, fluorophore-tagged SAF-A/hnRNP U is enriched at the Xi (Helbig and Fackelmayer 2003, Fackelmayer 2005, Pullirsch, Hartel et al. 2010). It also co-localizes with Xist RNA in conventional fluorescence microscopy imaging. These findings stand in stark contrast to the observations made with the antibody raised to the murine SAF-A/hnRNP U protein (see Chapter 2). Noteworthy, the relative enrichment of SAF-A/hnRNP U at the Xi appears to be unique to GFP-tagged constructs. Attempts to visualize the Xi using FLAG-tagged full length SAF-A/hnRNP U in Neuro2a cells were unsuccessful (Figure 4.5). Rather than highlighting the Xi, IF imaging on FLAG-tagged protein gives rise to a broad nuclear staining pattern that resembles the picture seen for endogenous SAF-A/hnRNP U. Since in both cases antibodies were required to visualize SAF-A/hnRNP U protein, this finding reinforces the idea that epitope occlusion is part of the Xi visualization issue. Furthermore, the possibility that the positioning of the tagging system itself may affect SAF-A/hnRNP U localization cannot be ignored. Previously it had been shown that there can be differences in protein stability, folding and function between N-and C-terminally tagged species of the same protein (Palmer and Freeman 2004). In case of GFP the tag localization did not appear to visibly influence SAF-A/hnRNP U localization, as in both the present and previous studies C-terminally hrGFP-tagged and N-terminally GFP-tagged SAF-A/hnRNP U were enriched at the Xi (Helbig and Fackelmayer 2003, Pullirsch, Hartel et al. 2010). FLAG-tagged protein, on the other hand, has not been studied in such detail. Therefore only data

on N-terminally tagged constructs showing no SAF-A/hnRNP U enrichment at the Xi have been conducted (here and in Hasegawa, Brockdorff et al. 2010).

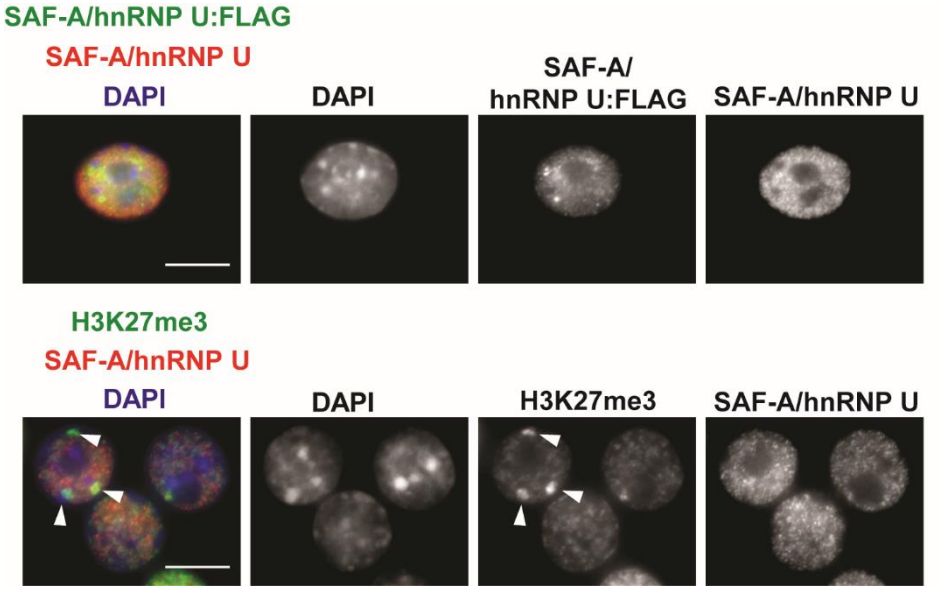


Figure 4.5 – Endogenous and FLAG-tagged SAF-A/hnRNP U in Neuro2a cells.

The two panels showing IF images of representative Neuro2a cells that stably express SAF-A/hnRNP U:FLAG. Images were taken on fixed cells that were immuno-stained with anti-FLAG (green in the merged image, upper panel), anti-H3K27me3 (green in the merged image, lower panel) and in house anti-SAF-A/hnRNP U (red in merged images of both panels). DNA was counterstained with DAPI (blue in merged images). Arrowheads (▼) point at Xis only visible in the H3K27me3 IF. Scale bars 10 µm.

When looking at the fusion protein, issue rational design would probably favor C-terminal tagging strategies for studies that focus on protein localization. This can be readily explained by looking at the domain structure of SAF-A/hnRNP U (Chapter 2, Figure 2.1) where the DNA binding domain is found at the N-terminus and the RNA binding domain on the C-terminus. This notion had been underpinned by deletion studies that indicate that the DNA

binding SAF box is essential for SAF-A/hnRNP U localization (Pullirsch, Hartel et al. 2010). In the future effects of the various tagging systems and their positioning on the N-or C-terminus need to be investigated further.

4.4.2 SAF-A/hnRNP U:hrGFP dynamics at the Xi and in the nuclear surroundings

To investigate SAF-A/hnRNP U dynamics in living cells, FRAP and iFRAP were used to specifically photobleach nuclear areas of interest (Figures 4.3 and 4.4). For the FRAP experiment different combinations of complete, incomplete and/or no recovery of fluorescence in the two bleached areas of interest give rise to nine possible outcomes for the experiment each with different biological implications. All possible combinations of events are summarized in the matrix in Figure 4.6A. The actual experiment showed a near complete recovery of the nuclear pool and only incomplete recovery dynamics at the Xi (Figures 4.3, 4.4 and 4.6).

These observations underpin the notion that SAF-A/hnRNP U exists in two distinct populations (see also Chapter 2, Figure 2.14). On the one hand, there is the SAF-A/hnRNP U population that stably associates with the Xi, and on the other hand there is the mobile protein fraction. The presence of the latter may account for the signal recovery up to non-Xi levels on the Xi. This means that mobile SAF-A/hnRNP U may be able to penetrate the Xi to a certain extent. However, it seems unlikely that the mobile fraction is involved in XCI. Immobile SAF-A/hnRNP U on the other hand could help to retain Xist RNA around the Xi territory. As mentioned earlier, SAF-A/hnRNP U is capable of self-assembling upon binding to DNA. Self-assembly is one of the potential mechanisms by which the protein achieves a strong association with the Xi (Fackelmayer, Dahm et al. 1994, Fackelmayer and Richter 1994). Importantly, this does not necessarily exclude any of the other mechanisms discussed in Chapter 2. Given the natural complexity of cellular processes it may be that they are all acting jointly or stand in a causative relationship to one another, i.e. a PTM may

induce conformational changes in the first place and facilitate protein multimerization thereafter.

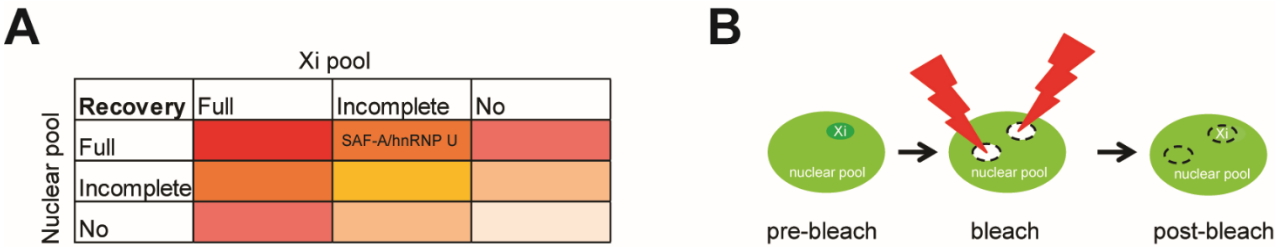


Figure 4.6 – SAF-A/hnRNP U FRAP summary.

(A) Table summarizing the nine (theoretically) possible outcomes for the SAF-A/hnRNP U FRAP experiment. (B) Schematics of the fluorescence recovery after photobleaching (FRAP) experiment that includes pre-bleach imaging, a bleaching step and imaging after photobleaching (post-bleach). Only the actual outcome of the SAF-A/hnRNP U FRAP experiment is shown: The nuclear pool fully recovers back to initial fluorescence levels and the Xi pool plateaus at the fluorescence level of the nuclear protein pool.

The apparent mobility of SAF-A/hnRNP U within the nuclear pool seen here also raises the question of whether there is actually a genuine association between SAF-A/hnRNP U and the nuclear matrix. The findings presented in this thesis support the conception that the nuclear matrix is a mere artifact created during a biochemical extraction procedure (Hancock 2000, Wilson and Coverley 2013).

5. SAF-A/hnRNP U does not associate with the Xi during mitosis

5.1 Introduction

Previously published evidence showed that human interphase XIST RNA tightly associates with the Xi. Intriguingly, the lncRNA is retained at the chromosome even after exposure to nuclear matrix extraction procedures that include treatment with high-salt buffers and nucleases (Berezney and Coffey 1977, Clemson, McNeil et al. 1996). Since nucleic acids are digested away during matrix preparations, it seems unlikely that the XIST-Xi association is DNA-mediated. Therefore, the idea that XIST RNA localization is regulated by protein factors is prevalent throughout the literature (Chan, Zhang et al. 2011, Hasegawa and Nakagawa 2011, Agostini, Cirillo et al. 2013). Human XIST has been shown to dissociate from the Xi upon entry into mitosis (Clemson, McNeil et al. 1996, Hall, Byron et al. 2009). In contrast, mouse Xist RNA remains associated with the inactive chromosome at least until metaphase, and vole Xist even until anaphase (Lee and Jaenisch 1997, Duthie, Nesterova et al. 1999). To further the current understanding of Xist RNA dynamics during the cell cycle, Ng *et al.* (2011) developed a tagging system that allows for the tracking of Xist RNA movement in live mouse ESCs (Ng, Daigle et al. 2011). Their study utilized a transgenic, inducible Xist system and found that, at interphase, a significant amount of time (~1.5 hours) is required for Xist RNA to spread over an entire chromosome. When cells entered mitosis, Xist RNA dissociated from the chromosome and re-appeared only after completion of cell division.

The suggestion that Xist RNA dissociates from the Xi during mitosis was reinforced by an additional study investigating the potential causes for changes in XIST localization during the cell cycle (Hall, Byron et al. 2009). Conditions under which XIST RNA was released at interphase or maintained binding throughout mitosis were tested. As a result, two inhibitors with opposing effects on XIST RNA localization had been identified (c.f. Figure 5.3): One of which was the Protein Phosphatase 1 (PP1) inhibitor tautomycin that facilitated the

interphase release of XIST from the chromosome. The other compound (hesperadin) was an Aurora Kinase B (AURKB) inhibitor that blocked the release of RNA from metaphase chromosomes. The authors conclude that chromosome condensation or other mechanical features of mitosis are not absolutely necessary for XIST RNA binding.

In light of these studies it was interesting to see whether there was a connection between Xist/XIST RNA and SAF-A/hnRNP U mitotic behavior. Earlier works on the topic demonstrated that HeLa cell SAF-A/hnRNP U turns over dynamically during the cell cycle, and is essential for faithful cell division (Ma, Matsunaga et al. 2011). A more detailed analysis in mouse and monkey cell lines revealed that metaphase SAF-A/hnRNP U mainly stays outside the chromosome territories with only a minor proportion of protein localizing to the chromosome arms and centromeric regions (Kukalev, Lobov et al. 2009). Notably, neither of the studies used cell lines bearing a prominent Xi, marked by an interphase accumulation of SAF-A/hnRNP U. Given the connection between SAF-A/hnRNP U and Xist, it was important to understand whether SAF-A/hnRNP U plays a role in determining Xist localization by looking at the relative localization of SAF-A/hnRNP U and Xist RNA in respect to one another, and at elucidating the dynamics of SAF-A/hnRNP U throughout the cell cycle.

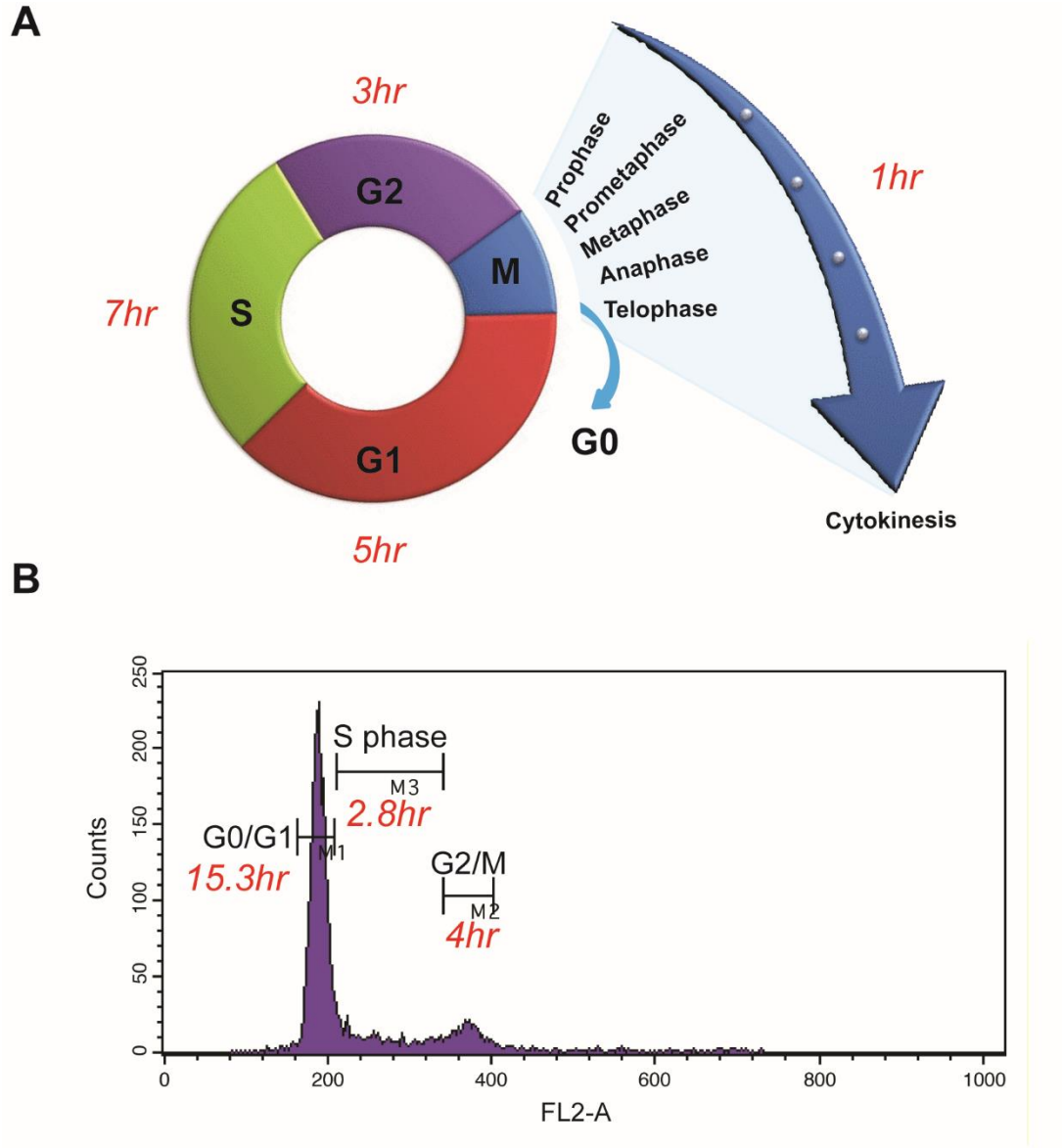
5.2 FSPE cell cycle analysis

Before performing critical experiments on the cell cycle dynamics of proteins in female *Mus spretus* ear tissue (FSPE) fibroblasts, which were the model system used in this set of experiments, a cell cycle analysis was performed using fluorescence-activated cell sorting (FACS). Figure 5.1A shows a standardized schematics of the mammalian cell cycle that can be divided into different stages: the gap phases G0, G1 and G2, a synthesis phase (S) at which the DNA is duplicated, and a mitotic phase (M) when two daughter cells arise from a single cell (Altinok, Gonze et al. 2011). Even though estimate timings are given for each of the cell cycle stages, the duration of individual steps may deviate significantly from the

numbers provided in Figure 5.1A. Such variability is usually due to the cell type, passage number and envrionmental factors. Observations on FSPE cells (n=10) in live cell imaging lead to the conclusion that an average cell cycle in these cells is completed in about 20 hours. In order to assess whether the cells undergo normal cell cycle progression, and to determine what proportion of cells was at a distinct cell cycle stage, several FACS runs with the DNA marker dye propidium iodide (PI) were conducted. Figure 5.1B shows a representative histogram plot and a table displaying the figures associated with this analysis. The table also provides information on the proportion of cells (%Gated) that were in either G0/G1,S phase or G2/M phase. These data were used to calculate the duration of the three distinct states using equation number (1) (Rabinovitch, Kubbies et al. 1988).

(1) $T(CS)=[T(CC)\times \ln(f(CS)+1)]/\ln 2$

Here T(CS) is the duration of the cell cycle stage, T(CC) the time it takes to complete an entire cell division cycle and f(CS) the fraction of cells at a particular stage as determined by FACS analysis. Notably, all computed numbers remain rough estimates that, in the future, may be used for guidance in the analysis of data obtained by FACS imaging (Chapter 5, Section 5.7) and for the purpose of cell synchronization experiments. Part B of Figure 5.1 also shows rounded numbers for the approximated duration of the three FACS-distinguishable phases. Due to the crude methodology used here, only additive values are available for G0/G1 and G2/M phases. Together G0 and G1 phase take about 15 hours, S phase 3 hours and G2/M phase roughly 4 hours.



Marker	Left, Right	Events	% Gated	% Total	Mean	Geo Mean
All	0, 1023	6155	100.00	75.57	227.77	217.32
M1	162, 207	4305	69.94	52.85	187.75	187.56
M2	340, 401	635	10.32	7.80	368.83	368.52
M3	210, 340	908	14.75	11.15	262.81	259.69

Figure 5.1 – Schematics of a typical eukaryotic cell cycle and FSPE cell cycle profile.

(A) Standard eukaryotic cell cycle divided into gap phase 1 (G1; red), synthesis phase (S; green), gap phase 2 (G2; violet) and mitosis (M; blue). The average timings of these are indicated in hours (*red*). Moreover, M phase can be subdivided into prophase, prometaphase, metaphase, anaphase and telophase. It is usually followed by cytokinesis after which the cell

remains either dormant, in gap phase 0 (G0), or proceed into G1. (B) Female *Mus spretus* ear tissue (FSPE) fibroblast cell cycle profile generated from fluorescence-activated cell sorting (FACS) data of propidium iodide (PI)-stained cells. In the plot G0/G1-, S- and G2/M-phases can be differentiated. Three gates were drawn to determine the proportion of cells at individual cell cycle stages. Numbers obtained were used to estimate the residency time of cells in the individual stages here indicated below the gates (*red*).

5.3 SAF-A/hnRNP U and Xist RNA cell cycle dynamics are different

Initially, to look at SAF-A/hnRNP U localization relative to Xist RNA in the course of a cell cycle, FSPE cells stably expressing SAF-A/hnRNP U:hrGFP (Chapter 4, Section 4.2) were subjected to Immuno-FISH treatment using a Xist RNA probe and a polyclonal hrGFP antibody (Figure 5.2). The specimen were examined with a fluorescence microscope, and cells at all mitotic stages and interphase were imaged. Examination of the individual stages showed that FSPE Xist remained associated with the Xi until prometaphase. No Xist RNA was seen on the chromosomes in meta-, ana- and telophase. Interestingly, SAF-A/hnRNP U appeared to be released from the Xi at the beginning of mitosis such that, already at prophase, Xist RNA and SAF-A/hnRNP U started dissociating from one another. At the prometaphase stage SAF-A/hnRNP U started to be excluded from the nucleus and returned at interphase only, when it was enriched at the Xi again.

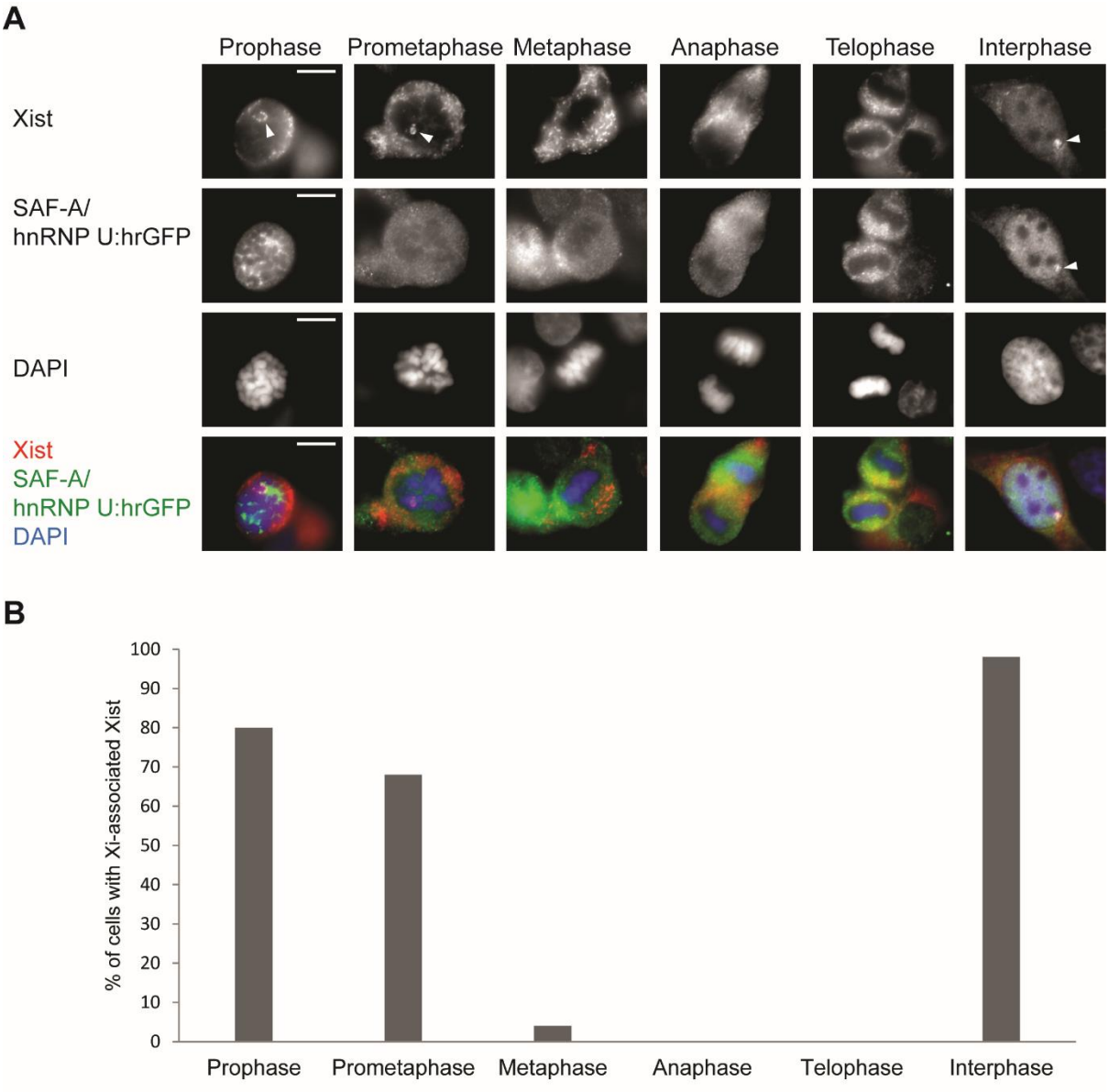


Figure 5.2 – Relative localization of SAF-A/hnRNP U and Xist RNA in FSPEs throughout the cell cycle.

(A) Exemplary immuno-FISH images on FSPEs with anti-hrGFP and Xist RNA probes. DNA was counterstained with DAPI. Representative images of individual microscopy channels (grayscale) and merged images are shown (color; Xist RNA – red, SAF-A/hnRNP U:hrGFP – green, DAPI – blue). Arrowheads (▼) pinpoint Xi-associated Xist RNA and SAF-A/hnRNP U in individual microscopy channels. Scale bar 10 µm. (B) Percentages of cell a with Xist foci at the Xi within each of the six cell cycle stages displayed here (n=50 per cell cycle stage analyzed).

After having assessed SAF-A/hnRNP U and Xist RNA turnover relative to one another it was interesting to see whether these dynamics can be perturbed using chemical agents that had previously been shown to alter Xist RNA localization patterns (Hall, Byron et al. 2009).

5.4 The effect of phosphatase and kinase inhibitors on SAF-A/hnRNP U localization

The previous section showed that SAF-A/hnRNP U and Xist RNA have distinct cell cycle dynamics in fixed cells. Whereas the protein is being released from the Xi at earlier stages of mitosis and recovers late after successful cell division, Xist RNA remains tightly associated with the Xi until the cell reaches prometaphase. This, however, does not mean that Xist localization is totally independent of SAF-A/hnRNP U because Xi-bound protein may still contribute to Xist RNA retention to the Xi exclusively during interphase.

In 2009, Hall *et al.* found conditions under which XIST RNA binding to the Xi was altered (Hall, Byron et al. 2009). If indeed SAF-A/hnRNP U was an Xist RNA retention factor at interphase, it may be that these conditions also change SAF-A/hnRNP U-Xi associations throughout the cell cycle. To answer this question the two drugs (tautomycin and hesperadin) used in the Hall *et al.* study were tested on FSPEs stably expressing SAF-A/hnRNP U:hrGFP (Figure 5.3).

Briefly, tautomycin was isolated in 1987 and shown to be a specific inhibitor of PP1 that is a serine/threonine phosphatase (Sydnes and Isobe 2010). PP1 can be found in all tissues and is one of the major enzymes that dephosphorylate serine/threonine residues in eukaryotic cells. For this reason it is involved in a wide range of cellular processes, i.e. protein synthesis, transcription and cell cycle progression. The small molecule inhibitor hesperadin, on the other hand, has been shown to inhibit AURKB that is a serine/threonine kinase involved in DNA condensation, chromosome orientation, kinetochore-microtubule

attachment and cytokinesis (Figure 5.3) (Carmena and Earnshaw 2003, Hauf, Cole et al. 2003, Agnese, Bazan et al. 2007).

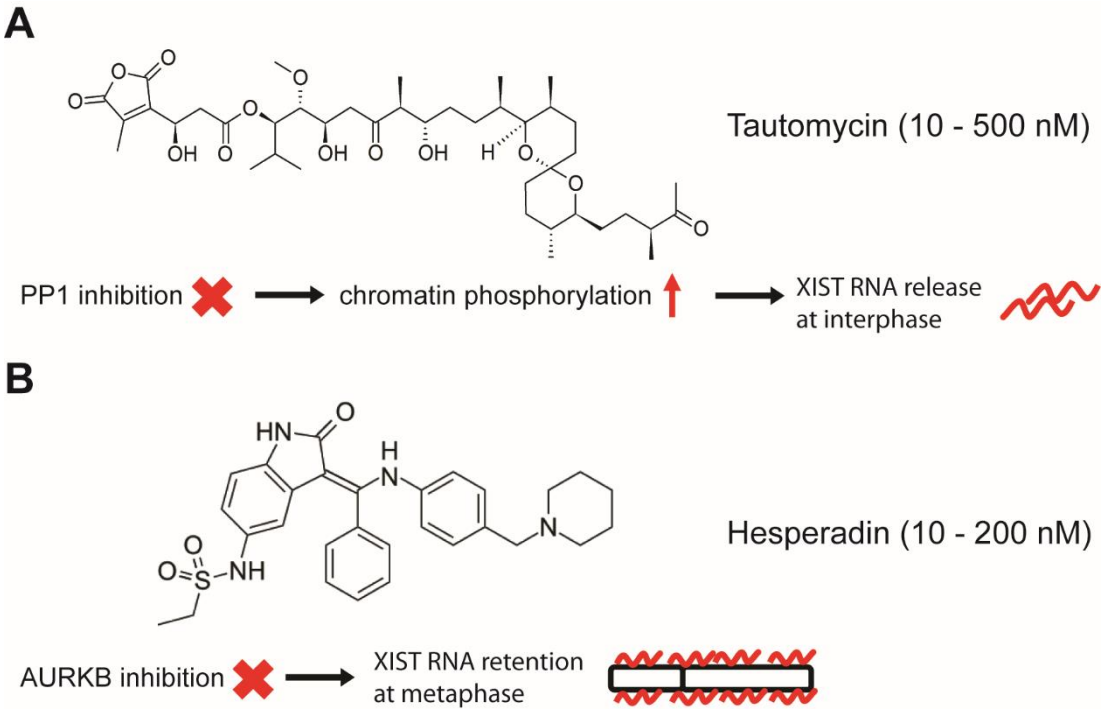


Figure 5.3 – Tautomycin and hesperadin: two drugs altering XIST-Xi associations.

(A) The protein phosphatase 1 (PP1) inhibitor molecule tautomycin increases overall chromatin phosphorylation, the effect of which is the ectopic XIST RNA dissociation from the interphase Xi (Hall, Byron et al. 2009). In this thesis work 10 – 500 nM tautomycin were used. (B) Hesperadin is an aurora kinase B (AURKB) inhibitor that had been shown to cause XIST retention to the Xi during metaphase (Hall, Byron et al. 2009). In this study a titration range of 10 – 200 nM hesperadin was tested on FSPE cells.

Both drugs were tested at various concentrations on FSPE cells in parallel with a dimethyl sulfoxide (DMSO) control. At the high end of the titration curve, which was at 200 – 500 nM tautomycin and 200 nM Hesperadin, cells showed altered morphology and cell death was observed (data not shown). Therefore, imaging data retrieved under these conditions were

disregarded in the subsequent analysis. Besides, due to the low number of cells at clearly identifiable, distinct cell cycle stages, all specimens were analyzed qualitatively only. Figure 5. 4 displays a microscopy panel with examples of interphase and mitotic cells treated with tautomycin, hesperadin and DMSO. For each of the drugs only two concentrations are shown. These were the two lowest concentrations used and are representative of the observations made even at near cytotoxic level. When looking at individual specimen, it became clear that tautomycin did not influence gross SAF-A/hnRNP U:hrGFP localization at interphase or mitosis (Figure 5. 4). Hesperadin treatment, on the contrary, seemed to result in a slight retention of SAF-A/hnRNP U at all chromosomes. Furthermore, a proportion of cells in prometaphase maintained SAF-A/hnRNP U:hrGFP enrichment at the Xi. Finally, as expected, exposure to the polar, aprotic solvent DMSO did not have any effects on SAF-A/hnRNP U:hrGFP sub-nuclear localization patterns.

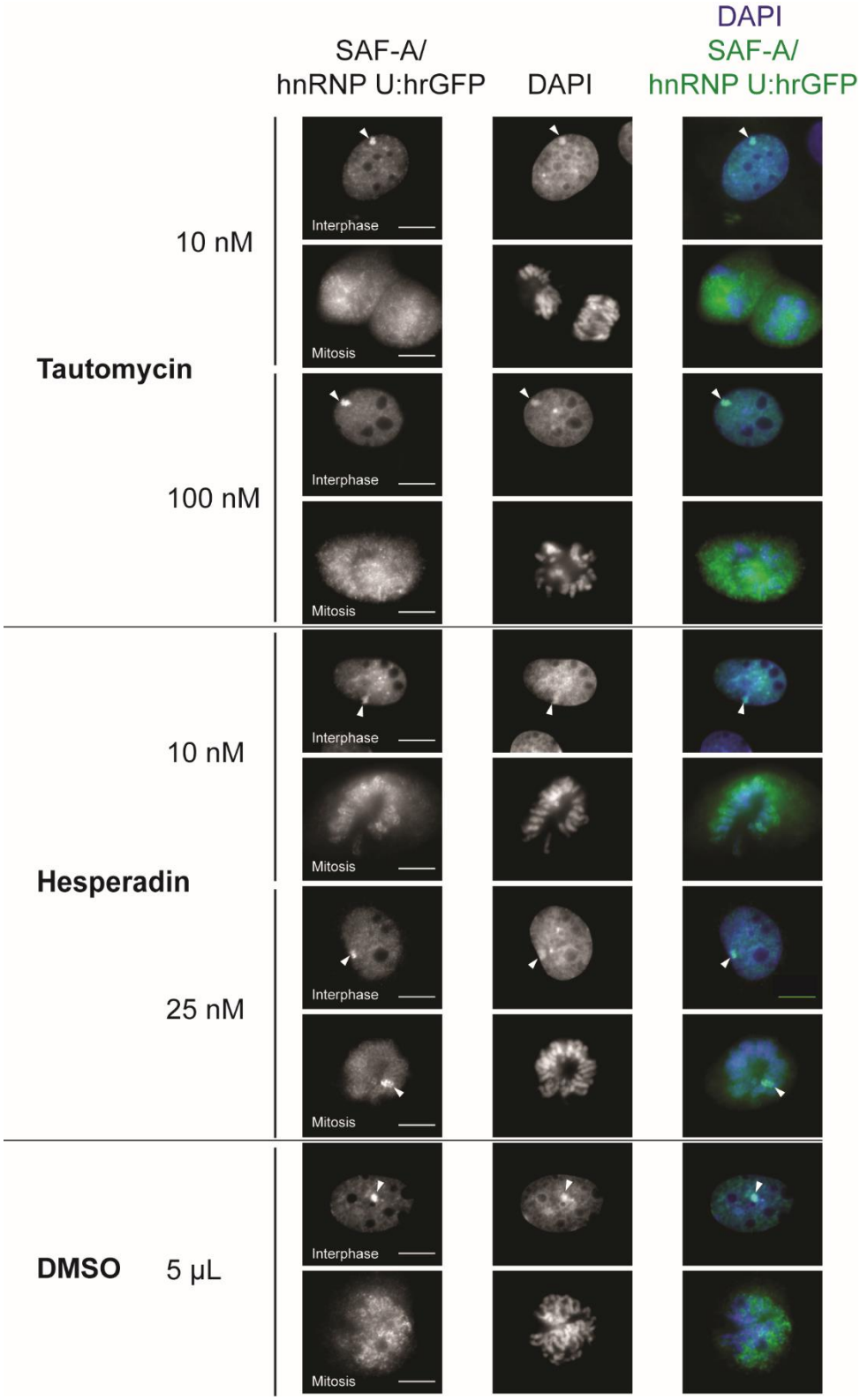


Figure 5. 4 – Drug treatment and SAF-A/hnRNP U localization in FSPE cells.

Representative images for FSPEs stably transfected with SAF-A/hnRNP U:hrGFP after 6 hours treatment with tautomycin (10 nM and 100 nM), hesperadin (10 nM and 25 nM) or 5 µL dimethyl

sulfoxide (DMSO). Cells were formaldehyde-fixed and DNA was counterstained with DAPI. Individual channels are shown in greyscale and merged images are colored with DAPI in blue and SAF-A/hnRNP U:hrGFP in green. For each condition representative interphase and mitotic cells are displayed. Mitotic cells are not strictly cell cycle sorted. Arrowheads (▼) pinpoint the Xi whenever appropriate. Scale bars 10 µm.

5.5 SAF-A/hnRNP U *in vivo* cell cycle dynamics

Because fixed cell imaging cannot capture the full dynamic range of the cell cycle and different cells are compared with one another, it is not the ideal technique to observe the transient nature of SAF-A/hnRNP U release and re-association with the Xi. Therefore, it was important to investigate SAF-A/hnRNP U dynamics *in vivo*. For this purpose SAF-A/hnRNP U live-cell dynamics were examined in a FSPE cell line that expressed hrGFP-tagged SAF-A/hnRNP U as well as the general chromatin marker H2B:mCherry (Figure 5.5 and Figure 5.6B).

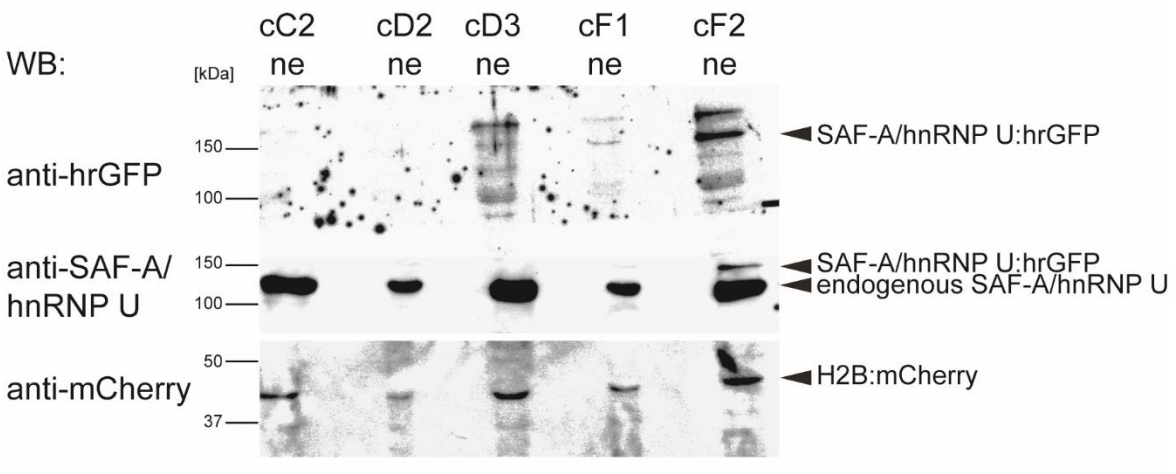


Figure 5.5 – Western blotting screen of SAF-A/hnRNP U:hrGFP – H2B:mCherry stable FSPE cell lines.

SAF-A/hnRNP U:hrGFP – H2B:mCherry positive cells were selected and grown up based on an initial fluorescence microscopy screen. However, the majority of clones did not express SAF-A/hnRNP U:hrGFP to a level detectable in Western blotting (WB). Here exemplary blots for five clones (cC2, cD2, cD3, cF1 and cF2), that were picked based on their appearance under the fluorescence microscope, are shown. Blots on equivalent amounts of nuclear extract (ne) were probed for hrGFP, SAF-A/hnRNP U and mCherry (top to bottom). Note that the hrGFP-tagged version of SAF-A/hnRNP U was clearly detectable in cF2 only. Bands corresponding to the individual proteins are indicated by arrowheads (▼).

Under the microscope individual cells were selected and followed over a time span of up to 24 hours (Figure 5.6A and C, respectively). After imaging, time-lapse movies were analyzed with a focus on dividing cells and the appearance/disappearance of the SAF-A/hnRNP U Xi spot (Figure 5.6C). As in fixed cells, SAF-A/hnRNP U:hrGFP was released at the beginning of mitosis (prophase) and did not recover until the cell division cycle had been completed. Once a cell had entered mitosis SAF-A/hnRNP U:hrGFP was excluded from the chromosome territories, and the majority of protein remained outside the nucleus even after nuclear envelope formation. SAF-A/hnRNP U enrichment at the Xi was detectable again late in G1 only.

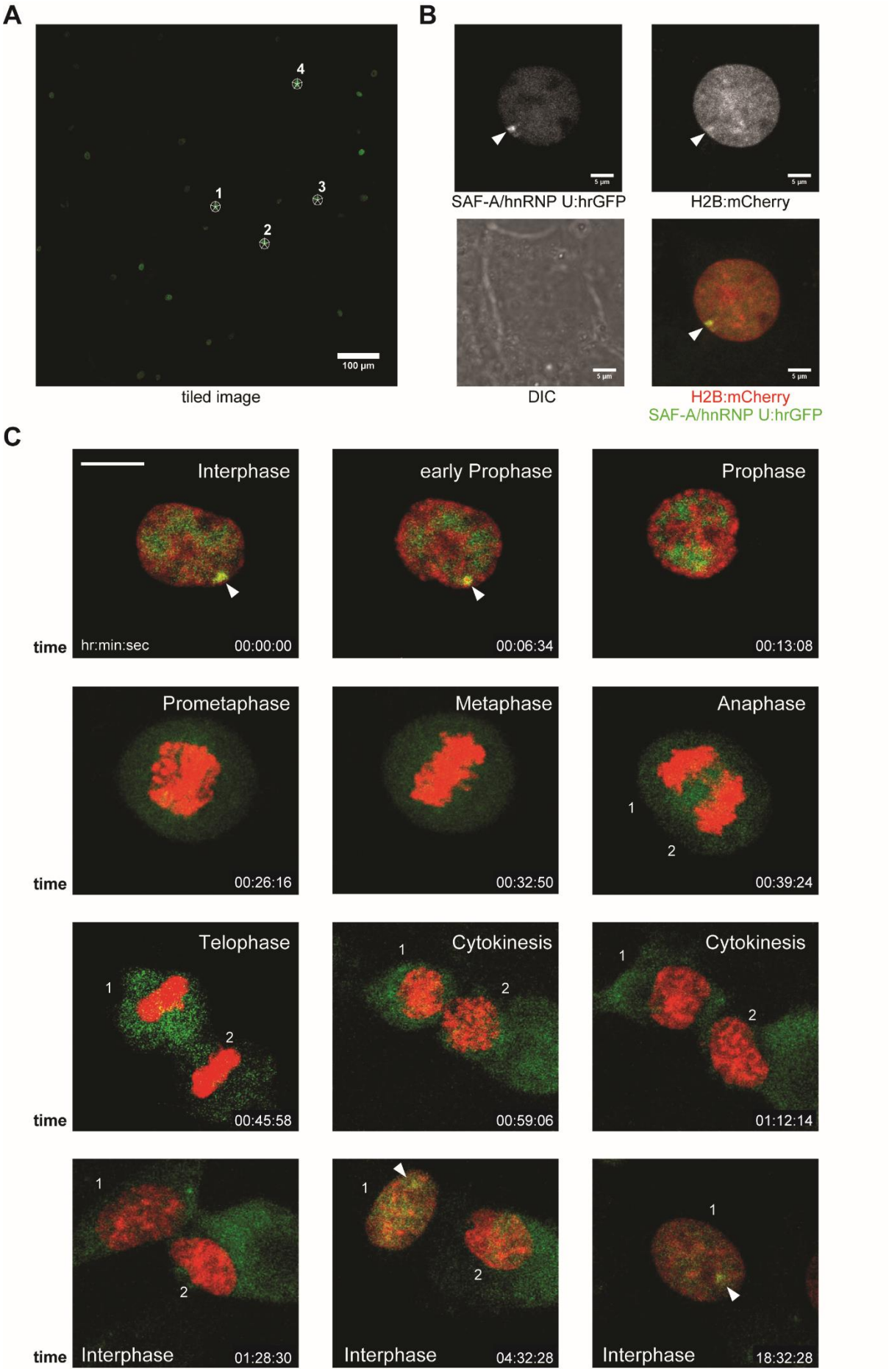

A
4
1
3
2
100 µm
tiled image
B
SAF-A/hnRNP U:hrGFP
H2B:mCherry
DIC
H2B:mCherry
SAF-A/hnRNP U:hrGFP
C
Interphase
early Prophase
Prophase
time
hr:min:sec
00:00:00
00:06:34
00:13:08
Prometaphase
Metaphase
Anaphase
time
00:26:16
00:32:50
00:39:24
Telophase
Cytokinesis
Cytokinesis
time
00:45:58
00:59:06
01:12:14
time
Interphase
01:28:30
Interphase
04:32:28
Interphase
18:32:28

Figure 5.6 – SAF-A/hnRNP U dynamics in single cell time-lapse imaging.

Single cell, tracked life-cell imaging of SAF-A/hnRNP U:hrGFP and H2B:mCherry stably expressed in FSPE cells. Cells were tracked using an European Molecular Biology Laboratory (EMBL)-developed macro (Rabut and Ellenberg 2004). (A) Example of a tiled image taken for initial cell selection. Four to five individual cells were followed at a time. Scale bar 100 µm. (B) Images of an individual interphase cell showing the different imaging channels in greyscale and a colored, merged image for SAF-A/hnRNP U:hrGFP (green) and H2B:mCherry (red). Arrowheads (▼) point at the Xi visible in two of the channels. In addition, a differential interference contrast (DIC) image is shown. Scale bars 5 µm. (C) Selected time-lapse images of SAF-A/hnRNP U (green) and H2B:mCherry (red). Relative imaging times are shown in a digital format (hr:min:sec) and cell cycle stages have been assigned to individual time points. For clarity, each of the two daughter cells had been assigned a number (one and two) upon initial chromosome segregation. In total 8 cells showing the same protein dynamics were tracked successfully using this method. Scale bar 10 µm.

The early dissociation from and late re-association of SAF-A/hnRNP U back to the Xi is a novel and interesting observation leaving open the question of whether there are other Xi-enriched proteins that show a similar behavior in the same cellular model system.

5.6 PCGF3 and SAF-A/hnRNP U cell cycle dynamics are similar

Recent examples in the Brockdorff laboratory have shown that in male XY, Xist-inducible ESCs several different GFP-tagged Polycomb Repressive Complex 1 (PRC1) proteins are recruited to the Xi upon Xist up-regulation (Mafalda Almeida, personal communication). Since ESCs are a model for initial Xist-dependent XCI, an important question to ask was whether the PRC1 components would behave similarly in a differentiated, XCI maintenance model system. Therefore, stable FSPE cell lines for N-terminally GFP-tagged Polycomb

Group Ring Finger Protein 1/3 (PCGF1:NT-GFP and PCGF3:NT-GFP) and N/C-terminally GFP-tagged Ring and YY1-binding protein (RYBP:NT-GFP and RYBP:CT-GFP) were prepared (Figure 5.7). All of these proteins are direct interaction partners of the E3 ubiquitin ligase Really Interesting New Gene 1B (RING1B) and were shown to specifically form variant PRC1 complexes (Farcas, Blackledge et al. 2012). Interestingly, in fixed cells only PCGF3:NT-GFP was clearly enriched at the Xi. The other cell lines bearing PCGF1:NT-GFP, RYBP:NT-GFP and RYBP:CT-GFP transgenes merely showed nuclear expression of the fluorophore-tagged protein products. To exclude that Xi-enrichment or masking was an artefact of the formaldehyde fixation procedure, all lines were examined in live-cell imaging. The observations were entirely consistent (Figure 5.7).

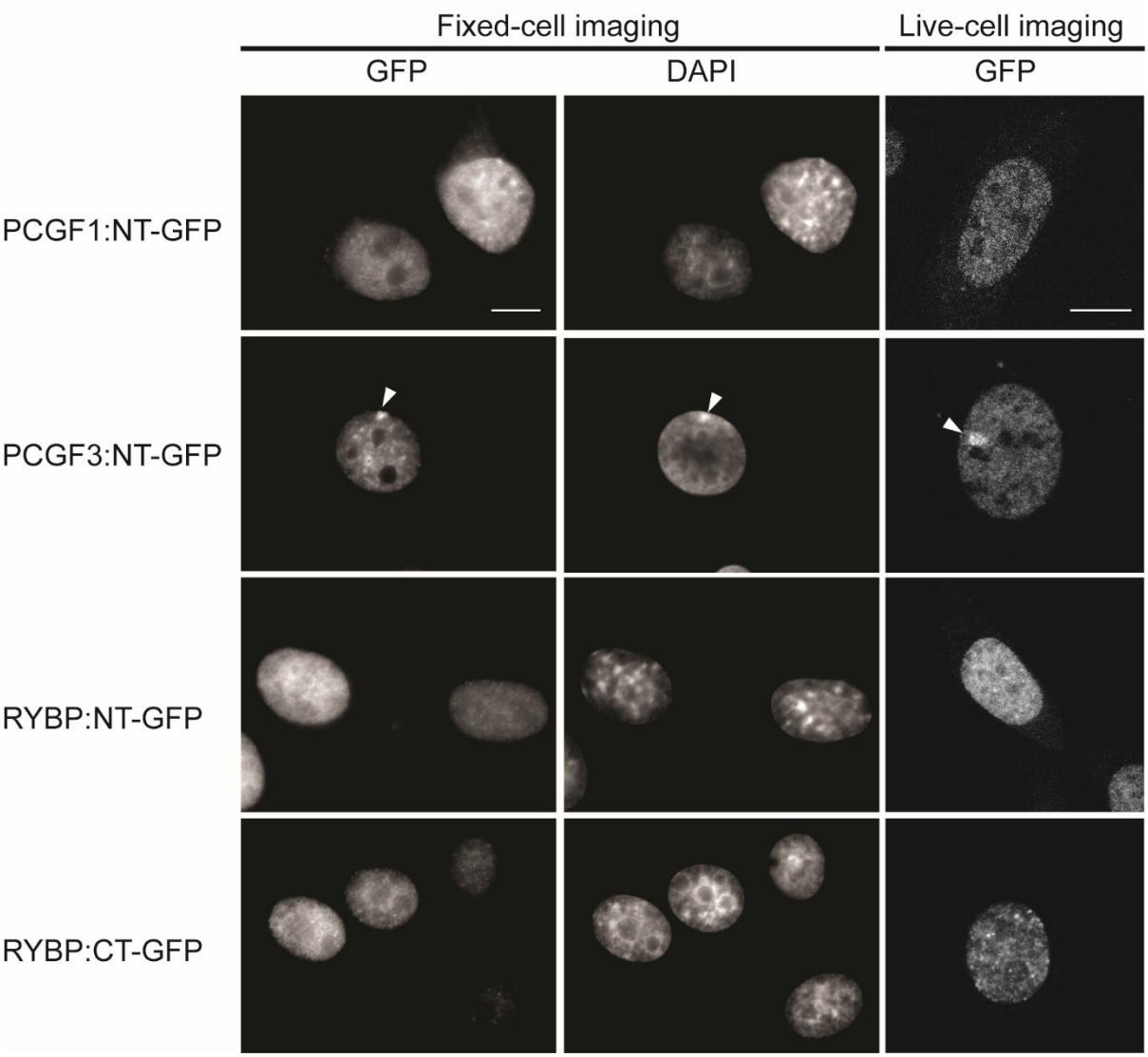


Figure 5.7 – Screen for Xi-association of PRC1 proteins in FSPE cells in fixed – and live – cell imaging.

FSPE cell lines expressing PCGF1, PCGF3 and RYBP bearing an N-terminal GFP-tag (NT-GFP), and C-terminally tagged RYBP were prepared and screened by conventional fluorescence microscopy (fixed cells) and confocal microscopy (live cells). For fixed cells, both GFP and DAPI channels are shown. The live-cell imaging panel displays the GFP channel only. All images are in greyscale. Scale bars represent 10 µm.

To further investigate its nuclear dynamics during the cell cycle, the PCGF3 lines were stably transfected with H2B:mCherry providing a chromatin marker (Figure 5.8A). Similar to SAF-A/hnRNP U, PCGF3 was released from the Xi once the cell started dividing, and only after completion of nuclear envelope formation was it recruited back to the inactive chromosome.

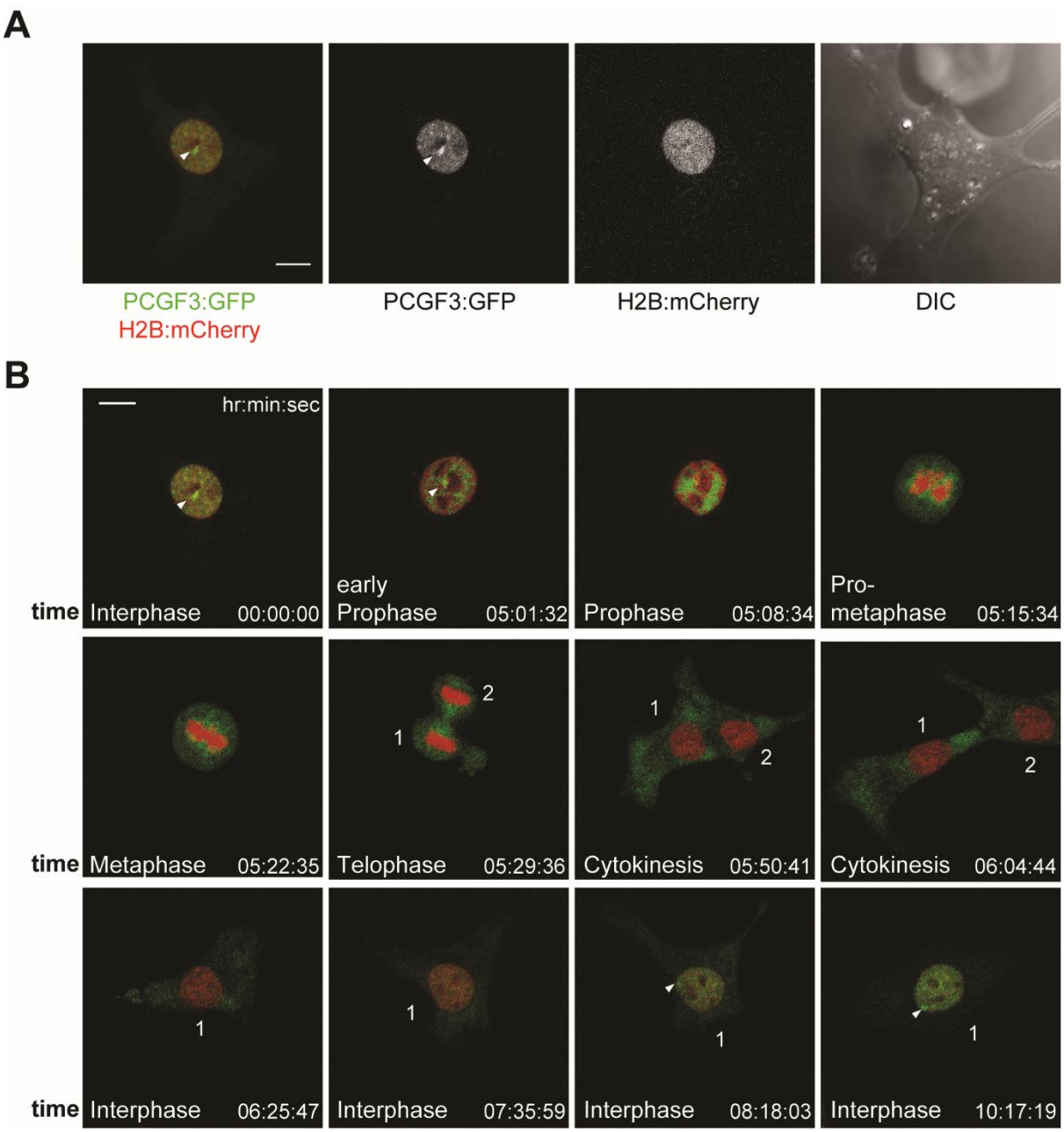


Figure 5.8 – PCGF3 dynamics in single cell, time-lapse imaging.

A single FSPE cell stably expressing PCGF3:GFP and H2B:mCherry was tracked and imaged throughout the cell cycle. Once again the EMBL tracking macro was used. (A) Images of an individual interphase cell showing the different imaging channels in greyscale, a merged image for PCGF3:GFP (green) and H2B:mCherry (red). Arrowheads (▼) point at the Xi. In addition a DIC image is shown. Scale bar 10 µm. (B) Selected time-lapse images showing PCGF3:GFP (green) and H2B:mCherry (red). Relative imaging times are displayed in digital format (hr:min:sec). Cell cycle stages have been assigned to individual time points. From telophase onwards the two daughter cells have been allocated numbers (one and two) to more easily differentiate the two. In total 3 cells were followed through a full cell division cycle.Scale bar 10 µm.

5.7 Preliminary FACS cell cycle and imaging analysis on SAF-A/hnRNP U and PCGF3

Since single-cell, time-lapse imaging proved to be technically difficult, alternative approaches were explored as well. FACS with an imaging flow cytometer allows for cell cycle sorting at the population level whilst taking snapshots of individual cells. Therefore, it was the method of choice used to complement the live-cell imaging data presented in Sections 5.5 and 5.6. Preliminary experiments comprising the initial optimization of the technique suggest that this approach maybe suitable for simultaneous analysis of cell cycle dynamics at the population and the single cell level.

Prior to imaging, cells stably expressing PCGF3:GFP, SAF-A/hnRNP U:hrGFP and appropriate control cell lines, i.e. the two lines without DAPI and the parental cell line with and without DAPI staining, were ethanol fixed and sample DNA was counterstained with DAPI. Thereafter all samples and controls were run on an Image Stream IS100 (Amnis) flow cytometer and the data were analyzed using the manufacturer's IDEAS software. Subsequent data analysis aimed at quantitatively analyzing the cell cycle-dependency of

disappearance and re-association of Xi-associated protein. Given the complexity of the system, this particular type of evaluation proved to be non-trivial. A closer look at the raw data in the Amnis IDEAS program clearly indicated that the images of the SAF-A/hnRNP U:hrGFP samples were of insufficient quality (inconsistency in fluorescence intensities in DAPI and hrGFP channels) to devise a general work flow for the analysis at this point. Hence, only individual pictures of an interphase FSPE with SAF-A/hnRNP U are shown in Figure 5.9.

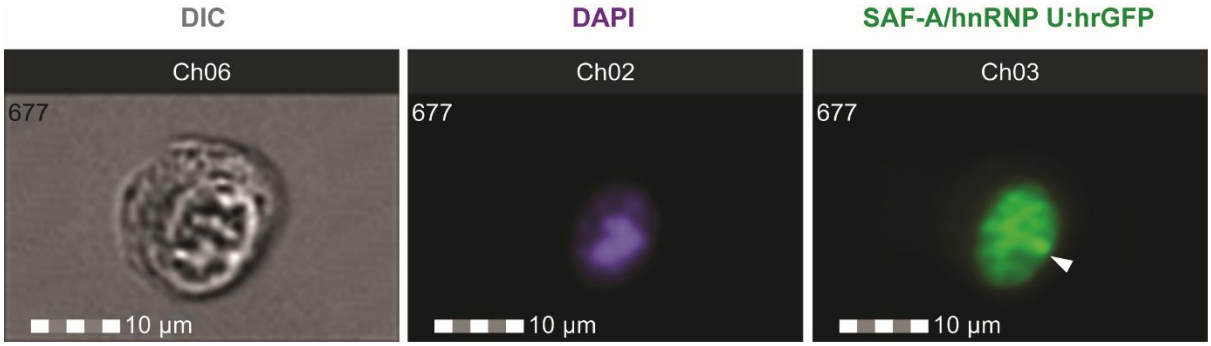


Figure 5.9 – FACS images of interphase SAF-A/hnRNP U:hrGFP.

Images taken in DIC (greyscale), DAPI (violet) and SAF-A/hnRNP U:hrGFP (green) channels. Scale bars 10 µm.

PCGF3:GFP data, however, proved to be of sufficient quality to attempt an initial analysis, the workflow of which is outlined in Figure 5.10. The FACS analysis workflow comprises an initial cell cycle stage sorting step followed by mask-based differentiation between cells with a single Xi-focus indicating PCGF3:GFP enrichment at this point (R14_1 GFP spot) and cells without focal PCGF3:GFP enrichment (R14_no spots). FACS plot number (9) in Figure 5.10 indicates that the current gating strategy is suboptimal as a clear-cut differentiation between the two pools could not be achieved. This observation, together with the high intensity of the

DAPI signal causing the histogram plot (4) to be skewed, impede a reliable quantitative analysis.

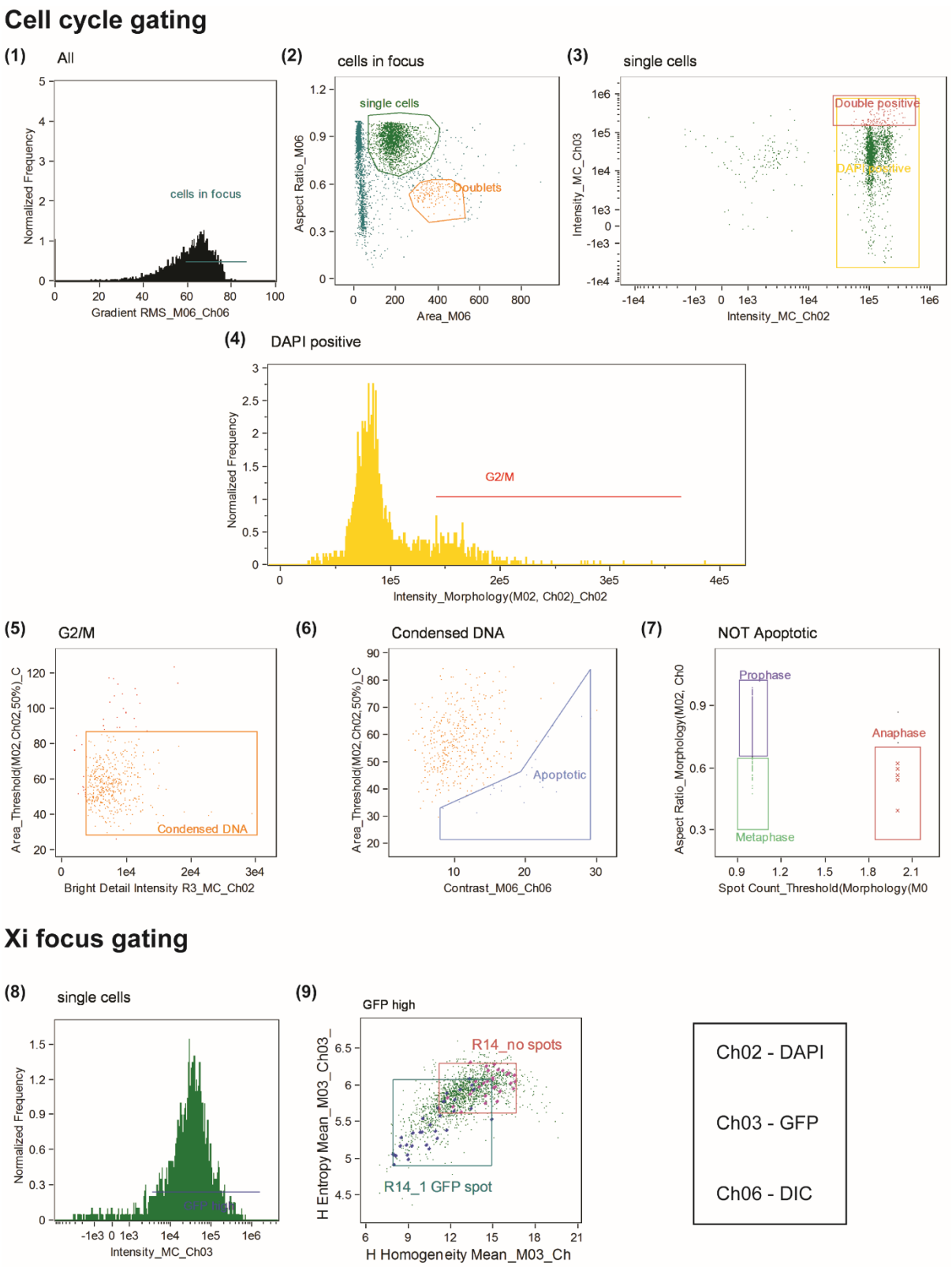


Figure 5.10 – Preliminary FACS gating strategies for cell cycle stages and PCGF3 at the Xi.

<u>Cell cycle gating:</u> (1) Firstly, all out of focus events were removed from the pool of imaged cells to be analyzed utilizing the DIC channel (Ch06). (2) Secondly, singlets, doublets and

clusters of cells were separated based on the diameter of the recorded events. (3) Subsequently, double positive events with high GFP and DAPI intensities were gated out. (4) These events were used to prepare a typical cell cycle histogram plot. Due to problems with fluorophore brightness and the current gating strategy the plot is an imperfect representation (compare with Figure 5.1). (5) To further narrow down the analysis a gate was drawn around DAPI-enriched G2- and M-phase cells. (6) Next apoptotic cells were subtracted. (7) Non-apoptotic cells were gated on three mitotic stages (prophase/metaphase/anaphase). Xi focus gating: (8) GFP positive single cells were gated out. (9) An attempt to differentiate between cells with a Xi and without Xi focus (R14_1 GFP spot and R14_no spots, respectively) was made utilizing a pre-defined GFP fluorescence intensity mask.

FACS imaging confirmed the results obtained in previous live-cell microscopy experiments (Section 5.5). PCGF3:GFP is enriched at the Xi during interphase and dissociates from the chromosome during mitosis (Figure 5.11). But due to the exploratory nature of the methodology no quantitative data could be derived from this pilot experimental run.

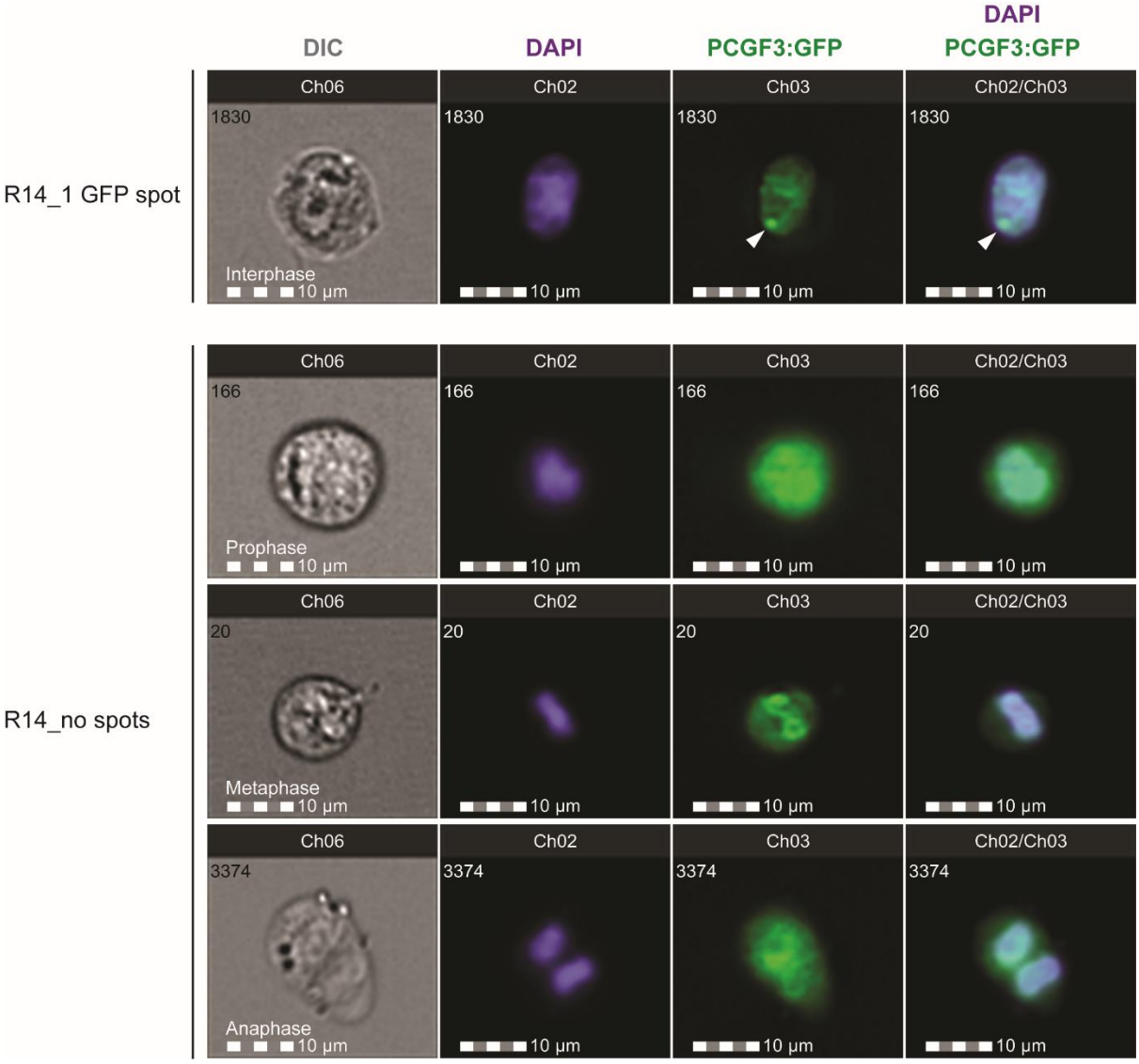


Figure 5.11 – FACS images of cell cycle sorted PCGF3:GFP events.

Individual cells were selected from the pools identified in Figure 5.10. DIC is shown in greyscale, individual and merged images in color with PCGF3:GFP in green and DAPI in violet. Most R14_1 GFP spot cells looked like the selected examples with a clear enrichment of PCGF3:GFP at the Xi (arrowhead, ▼). The R14_no spots population was more heterogeneous. By combining cell cycle and Xi focus gating analysis three mitotic stages could be differentiated. Displayed are images of three representative cells at pro-, meta- and anaphase. Notably, not all cells within the gated areas matched the indicated stage of mitosis (data not shown). Numbers in the left hand corner of the images are unique identifiers for individual cells acquired in the experiment. In some instances, i.e. the bottom most panel, the DIC image

displays small black dots that correspond to the beads used in the FACS machine setup. Scale bars 10 µm.

5.8 Summary and discussion

5.8.1 SAF-A/hnRNP U release precedes Xist RNA dissociation from the Xi

Earlier studies reporting on a direct interaction between SAF-A/hnRNP U and Xist RNA prompted the possibility that fluorophore-tagged versions of the protein could be used instead of the MS2-tagging system for Xist RNA developed by *Ng et al.* (Ng, Daigle et al. 2011). The MS2-tagging strategy is based on high affinity interactions between a viral RNA loop structure and the MS2 protein that allowed for the tracking of loop-tagged Xist RNA through the binding of GFP-tagged MS2 in live-cell microscopy experiments. Even though the principle of this methodology is very elegant, it comes with the drawback that cell viability and vitality are compromised (personal communication with Prof. Dr. Neil Brockdorff). Therefore, instead of introducing an exogenous phage protein and undertaking the effort of tagging Xist RNA, having a traceable, natural Xist RNA binder would be of great use in studies further investigating Xist/XIST RNA dynamics in living cells. To investigate the usefulness of SAF-A/hnRNP U in such a setup, initial tests needed to be run. First of all SAF-A/hnRNP U and Xist localization in respect to another, at all stages of the cell cycle, was assessed in SAF-A/hnRNP U:hrGFP immuno-Xist RNA-FISH experiments. These revealed that the dynamics of the two macromolecules are markedly distinct. Consequently it may be assumed that, if there was a direct association between the two, it is either transient, unstable or cell-cycle regulated. The observation that SAF-A/hnRNP U clearly dissociates from the Xi earlier than Xist RNA potentially indicates that there are no direct interactions between the two. This conclusion contradicts published literature that provided evidence by UV-crosslinking RNA IP to show that SAF-A/hnRNP U directly interacts with Xist RNA exon 1 (Hasegawa, Brockdorff et al. 2010). Viewed from this angle Hasegawa *et al.*'s (2010)

observation could be an artefact of the IP experiment or the result of a more complex biological process, i.e. a SAF-A/hnRNP U association that exists only during interphase and detachment of the two macromolecules from one another upon entry to mitosis.

5.8.2 Influence of tautomycin and hesperadin on SAF-A/hnRNP U localization

In summary, the preliminary inhibitor studies suggest that SAF-A/hnRNP U accumulation at the Xi is independent of PP1 activity, and excludes the possibility that SAF-A/hnRNP U is a direct target of the phosphatase. Inhibition of its antagonist AURKB, however, resulted in longer retention times for SAF-A/hnRNP U at the Xi. Interestingly, it is believed that the opposing action of the two proteins is based on a common substrate recognition pattern (Posch, Khoudoli et al. 2010, Tatchell, Makrantoni et al. 2011). Therefore, the effect seen with the specific kinase inhibitor hesperadin is unlikely to be due to AURKB-mediated phosphorylation of SAF-A/hnRNP U. This together with the error tolerant Mascot analysis in Chapter 2 (Section 2.4), that failed to identify a previously published phosphorylation site on the protein, strongly suggests that the SAF-A/hnRNP U-Xi association is not regulated directly by protein phosphorylation (Britton, Froment et al. 2009).

Nevertheless, it is important to mention that AURKB impacts on chromatin structure and chromosome segregation behavior (Lesage, Qian et al. 2011). Thus, it is equally likely that SAF-A/hnRNP U is passively retained to the Xi chromatin through an XCI-independent mechanism. The same holds true for XIST RNA that maintained its association with the Xi during initial stages of mitosis in a study conducted by Hall *et al.* (Hall, Byron et al. 2009). Once again it should be mentioned that mouse and human may not be entirely comparable and that, in contrast to the initial study using these inhibitors on human cells, the present experiments were all done using murine cells. Furthermore, as with all drug-based studies, great care needs to be taken at the interpretation stage because pleiotropic effects may alter the significance of the results. This is an important aspect to consider when designing future

experiments. Furthermore, Xist-FISH type experiments would greatly complement the present experimental panel.

5.8.3 SAF-A/hnRNP U and PCGF3 cell cycle dynamics are comparable

In Chapter 4 it was demonstrated that fluorophore-tagged interphase SAF-A/hnRNP U stably associates with the Xi. Here this finding was complemented by long-term, time-lapse imaging over an entire cell cycle. To have a point of reference another Xi-enriched protein factor was subjected to the same type of analysis: The PRC1 complex component PCGF3 was the only additional clearly enriched factor found in the present set of experiments. Therefore, SAF-A/hnRNP U and PCGF3 live-cell dynamics were investigated. It was found that, like many other proteins, SAF-A/hnRNP U and PCGF3 turn over dynamically during the cell division cycle. This study shows that both SAF-A/hnRNP U and PCGF3 dissociate from the Xi at the onset of mitosis. As ubiquitination-mediated protein degradation is a common pathway to regulate protein turnover in the cell, this observation ties in nicely with the observed ubiquitination pattern of SAF-A/hnRNP U (Chapter 2, Section 2.4). Whether their comparable dissociation and association kinetcs are merely correlated or whether there is a functional link, i.e. through common complex formation, between SAF-A/hnRNP U and PCGF3 needs to be addressed in forthcoming studies. Further additions that should be made to the current findings are comparative, quantitative time courses on the cell cycle dynamics of the two proteins. For this purpose suffcient numbers of cells need to be imaged and all data have to be summarized in a common analysis format. To date this proved to be technically difficult due to the high mobility of FSPE cells and unaccountable variability in H2B:mCherry brightness. A possible solution to this would be FACS imaging, where a variety of markers can be used to label the DNA and/or enhance the signal itensity of fluorophores through antibody-staining. Moreover, it is a quick and easy to use technique allowing for the acquisition of high cell numbers.

6. Endogenous and hrGFP-tagged SAF-A/hnRNP U in super-resolution imaging

6.1 Introduction

Insight into sub-cellular organization has long been limited by the optical resolution power of the techniques used. Due to diffraction limits classical fluorescence microscopy has a spatial resolution of about 200 nm (Kner, Chhun et al. 2009). To go beyond this limit several methods have been developed, and each comes with its own technical constraints (Gustafsson 2000, Gustafsson 2005, Betzig, Patterson et al. 2006, Hess, Girirajan et al. 2006, Willig, Rizzoli et al. 2006). Localization-based methods like photoactivated localization microscopy (PALM), stochastic optical reconstruction microscopy (STORM) and fluorescence photoactivation localization microscopy (FPALM) can achieve high resolutions, but they also require large amounts of raw data that restrict the imaging speed. With stimulated emission depletion microscopy (STED), on the contrary, two points of less than 30 nm distance can be resolved at a high frame rate (Westphal, Rizzoli et al. 2008). This, however, comes with the drawback that the STED imaging field is very small. A technique that, in terms of performance, falls in between all the aforementioned categories is three-dimensional structured illumination microscopy (3D-SIM). 3D-SIM usually achieves resolutions of about 100 nm and allows imaging of larger areas than STED at any given frame rate (Markaki, Smeets et al. 2012). Figure 6.1A illustrates the 3D-SIM optical resolution by shifting directly overlaid images in 20 nm, 40 nm, 80 nm, 160 nm and 320 nm increments. It also shows that a visible separation of two points can be achieved already at a distance of 80 nm. To show the power of 3D-SIM even better, Figure 6.1B compares a nucleus in both wide-field deconvolution microscopy with 3D-SIM imaging.

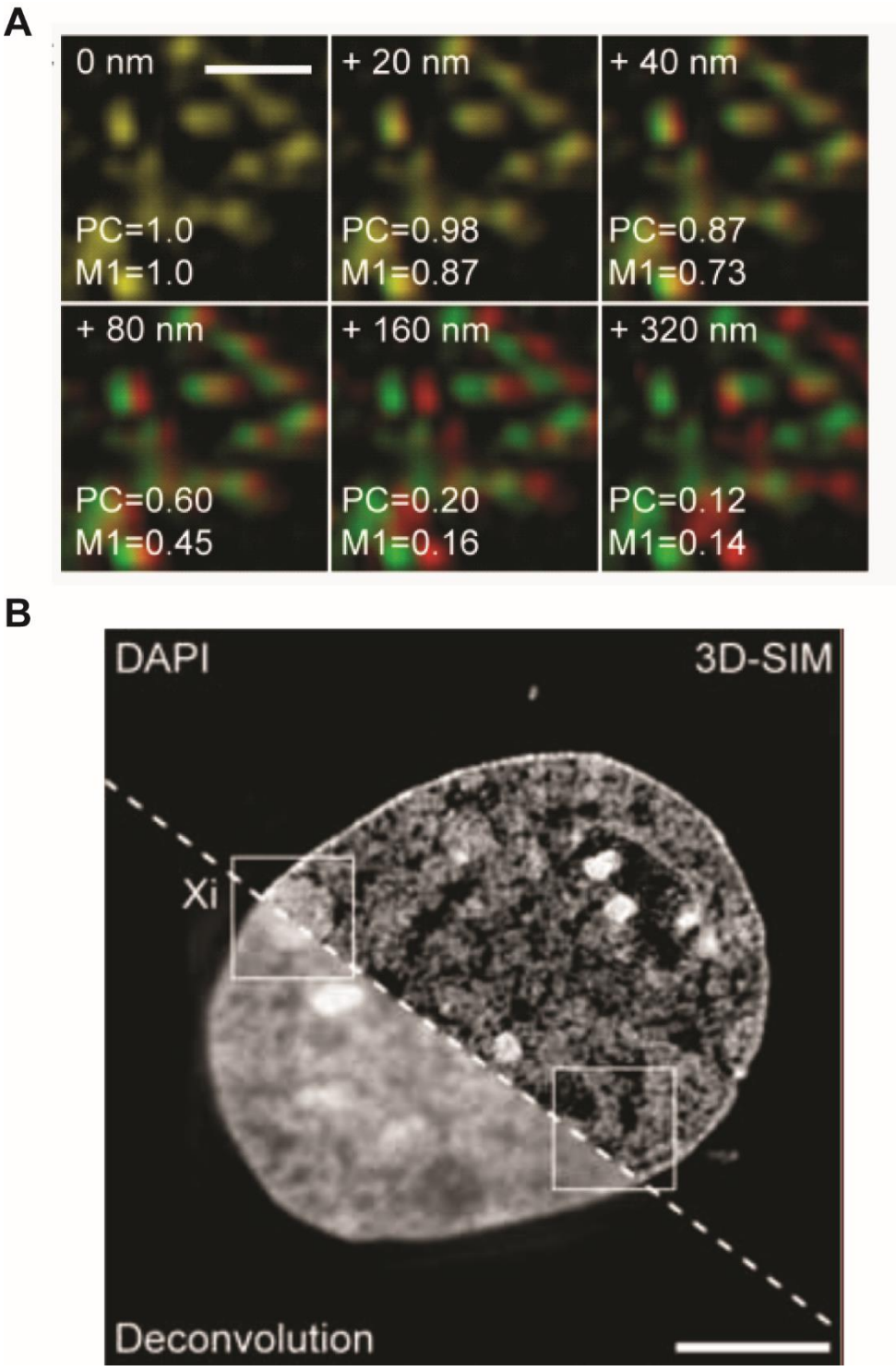


Figure 6.1 – Co-localization in and resolution of 3D-SIM.

(A) Exemplary co-localization values and absolute distances of an image that was split in two channels (green and red) *in silico*. Then red and green images were shifted such that the spots were separated 20 nm, 40 nm, 80 nm, 160 nm and 320 nm, respectively. The corresponding correlation coefficients (PC – Pearson's coefficient and M1 – Manders' 1 coefficient) and their

change with increasing spot separation are displayed as well. Scale bar 0.5 µm (by courtesy of Dr. Daniel Smeets, group of Dr. Lothar Schermelleh). (B) Comparison of DAPI-stained nucleus in wide-field microscopy vs. 3D-SIM (Smeets, Markaki et al. 2014). Scale bar 1 µm.

Briefly, in 3D-SIM the sample is illuminated with spatially structured excitation light that makes high-resolution information visible. By using moiré patterns, high-resolution reconstructions are prepared after initial data acquisition. Moiré fringes are the product of superimposition (multiplication) of fine grids (Figure 6.2) and the resulting patterns are used to computationally derive information on the original sample structures.

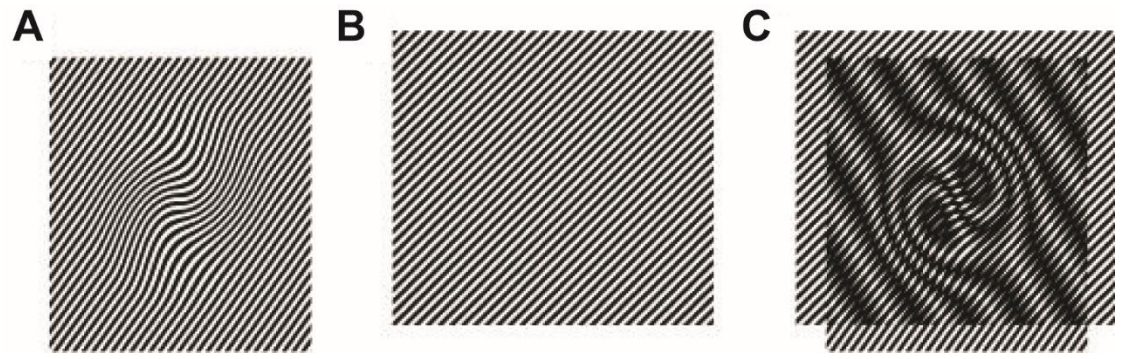


Figure 6.2 – Enhanced image resolution through the moiré effect.

(A) Line pattern of unknown sample structure. (B) Known regular illumination pattern. (C) Superimposed (multiplied) patterns showing moiré fringes seen as apparently darker structures. Sample information can be computationally reconstructed from the obtained pattern (Figure and legend were adapted from Gustafsson [2005] (Gustafsson 2005).).

The OMX is a microscope that can be used to perform SIM imaging. The instrument was designed and built by John Sedat and David Agard (University of California, San Francisco) with the support of a number of co-workers including Mats Gustafsson, Lukman Winoto and Peter Carlton (Goldman 2010). The name OMX stands for an optical microscope that is constantly evolving and therefore always at an “eXperimental” stage of development (Goldman 2010). For sample illumination the OMX uses solid-state, continuous-wave lasers to achieve bright, monochromatic illumination, and at least five lasers can be installed in an

OMX laser bed. These are bundled into two fibers that take alternative paths into the microscope body. One of these paths is used for fast live cell imaging, whilst the other one is used for 3D-SIM (Goldman 2010). As mentioned before, 3D-SIM uses structures within the illumination and multiple images per sample section that are combined with extensive data analysis to achieve double the resolution in all three dimensions, and an about eightfold improvement in volumetric resolution (Smeets, Markaki et al. 2014). By illuminating the fluorescent sample with striped, sinusoidal patterns, shifting these pattern in five steps, and rotating them to three positions by an 60˚angle over a full imaging cycle, the raw data for the 3D-SIM image are generated (Goldman 2010). Subsequently, all acquired images are processed using computer algorithms and a final high-resolution image of the specimen is generated (Schermelleh, Heintzmann et al. 2010). These cross-correlation and minimization algorithms use the acquired data to estimate experimental parameters such as grating constant, phase and direction, unmix components in the frequency space, and shift the moiré patterns in a way that facilitates the generation of a final image (Schermelleh, Heintzmann et al. 2010). To computationally reconstruct a high-resolution 3D image of a mammalian cell about 1,000 raw images need to be recorded. Additionally, the acquisition of control bead images is required to achieve proper image alignments (Schermelleh, Heintzmann et al. 2010).

In previous chapters (Chapters 2, 4 and 5) of this work SAF-A/hnRNP U sub-nuclear localization had been investigated. Even though hrGFP-tagged fusion protein clearly localized to the Xi in conventional fluorescence imaging, the resolution proved to be insufficient to draw any further conclusions. Therefore, endogenous and tagged SAF-A/hnRNP U were imaged at higher resolution with 3D-SIM. Furthermore, previously acquired data were complemented by comparing the sub-cellular localizations of mouse and human SAF-A/hnRNP U.

6.2 SAF-A/hnRNP U associates with Xist RNA

As shown in Chapter 2, the antibody raised against endogenous mouse SAF-A/hnRNP U did not show the protein's Xi-enrichment in conventional fluorescence microscopy. However, the resolution of conventional microscopy is limited. Therefore, another attempt to look at antibody-labelled SAF-A/hnRNP U in 3D-SIM was made. By doubling the resolution power it was hoped to gain a better understanding of SAF-A/hnRNP U localization within the nuclei of mouse cells in respect to Xist RNA. Here mouse myoblast cell line C2C12 were used. This cell line had been successfully used in a published study that amongst others also looked at the spatial relationship between Xist RNA and SAF-A/hnRNP U (Smeets, Markaki et al. 2014). Thus, data obtained in C2C12 cells could be compared to previously published evidence.

Figure 6.3A displays an exemplary z-section of three TetraSpeck bead images in different channels taken for instrument adjustment and alignment purposes in the later image analysis process. In addition, 3D-SIM images of a C127 (epithelial, *mus musculus*, mammary gland cells) test sample were taken and used in the later data alignment process (Figure 6.3B).

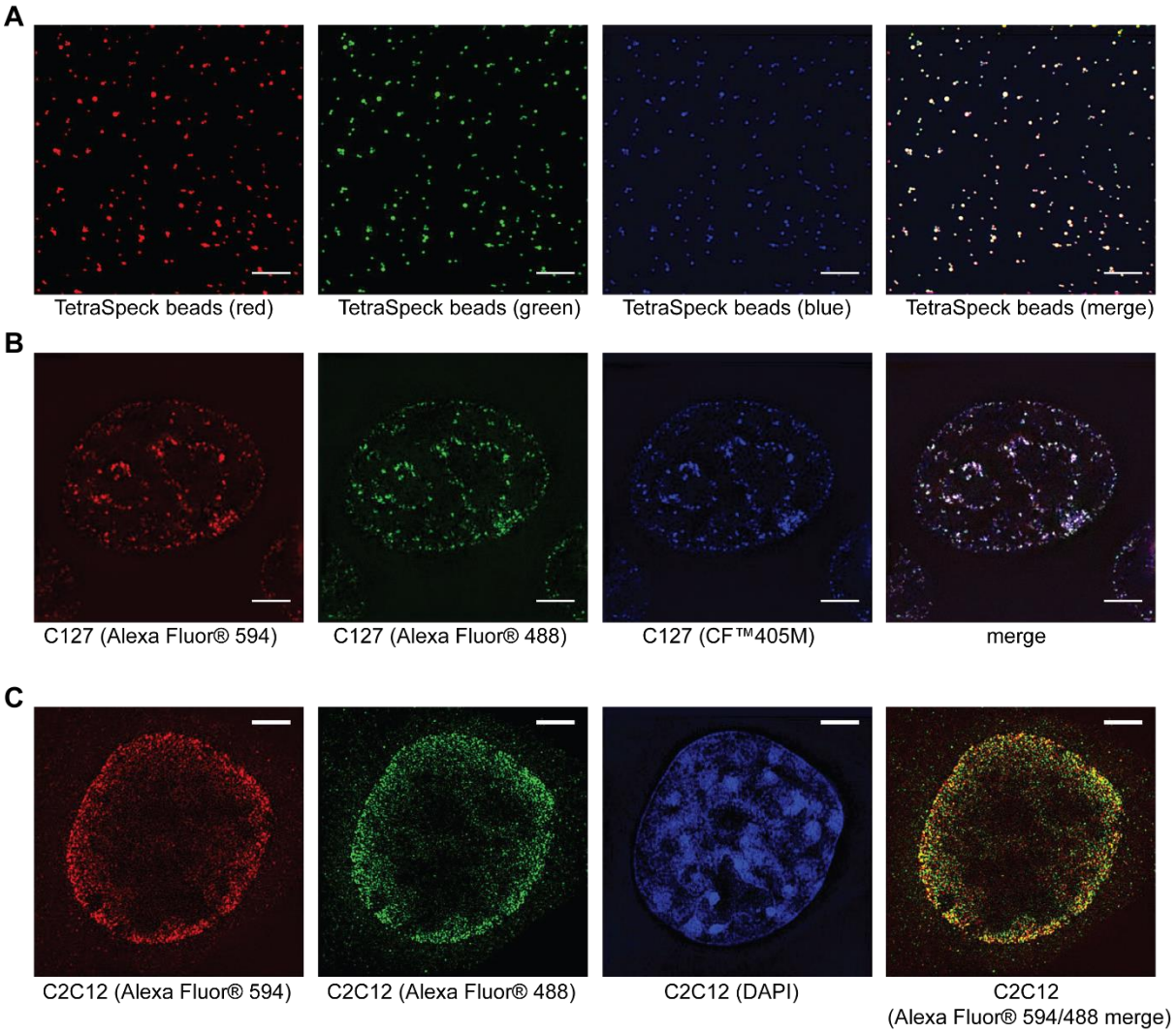


Figure 6.3 – 3D-SIM alignment and secondary antibody controls.

(A) TetraSpeck beads were imaged in three channels for alignment and optimization purposes. (Scale bars: 3 µm). (B) Images of a female C127 mouse mammary gland cell stained with three different fluorophores (Alexa Fluor® 594, 488 and CF™405M) taken for alignment purposes (Scale bars: 3 µm). (C) Images of C2C12 cells stained with SAF-A/hnRNP U primary antibody and two secondary antibodies coupled to two different fluorophores (Alexa Fluor® 594 and 488) (Scale bars: 3 µm).

A series of samples was prepared to examine the spatial relationship between Xist RNA and SAF-A/hnRNP U: Two specimen were subjected to anti-SAF-A/hnRNP U-Xist RNA FISH, with the colors used for each of the labels swapped between the samples. Another two

specimen were transiently transfected with SAF-A/hnRNP U constructs and subjected to anti-hrGFP-Xist RNA FISH, and once again the colors were swapped in between samples. Thirdly, an anti-SAF-A/hnRNP U control stained with two secondary antibodies of different color was prepared. Sample images of the latter one are displayed in Figure 6.3C and the yellow signal shows that both secondary antibodies had bound to epitopes in close proximity to another.

Even though the transfection of SAF-A/hnRNP U:hrGFP was conducted as described previously (Smeets, Markaki et al. 2014), the attempt to replicate the results failed because the transfected C2C12 cells did not express nuclear SAF-A/hnRNP U:hrGFP that was clearly enriched at the Xi (data not shown). Instead, a broad cytoplasmic staining pattern was observed in most of the samples. Therefore, all SAF-A/hnRNP U:hrGFP-transfected samples were excluded from this analysis. The antibody staining of the protein, on the contrary, was successful and further analysis (nearest neighbor distance determination) could be performed on the raw data. An overview of the data analysis workflow is displayed in Figure 6.4A, and Figure 6.4B illustrates this workflow on a practical example taken from this study. Aligned and structured illumination (SI)-reconstructed images of C2C12 cells were used to determine median distances between Xist RNA and SAF-A/hnRNP U centroids. These may provide information on the spatial relationship of the two macromolecules and the results of this analysis are displayed in Figures 6.5C and 6.5E.

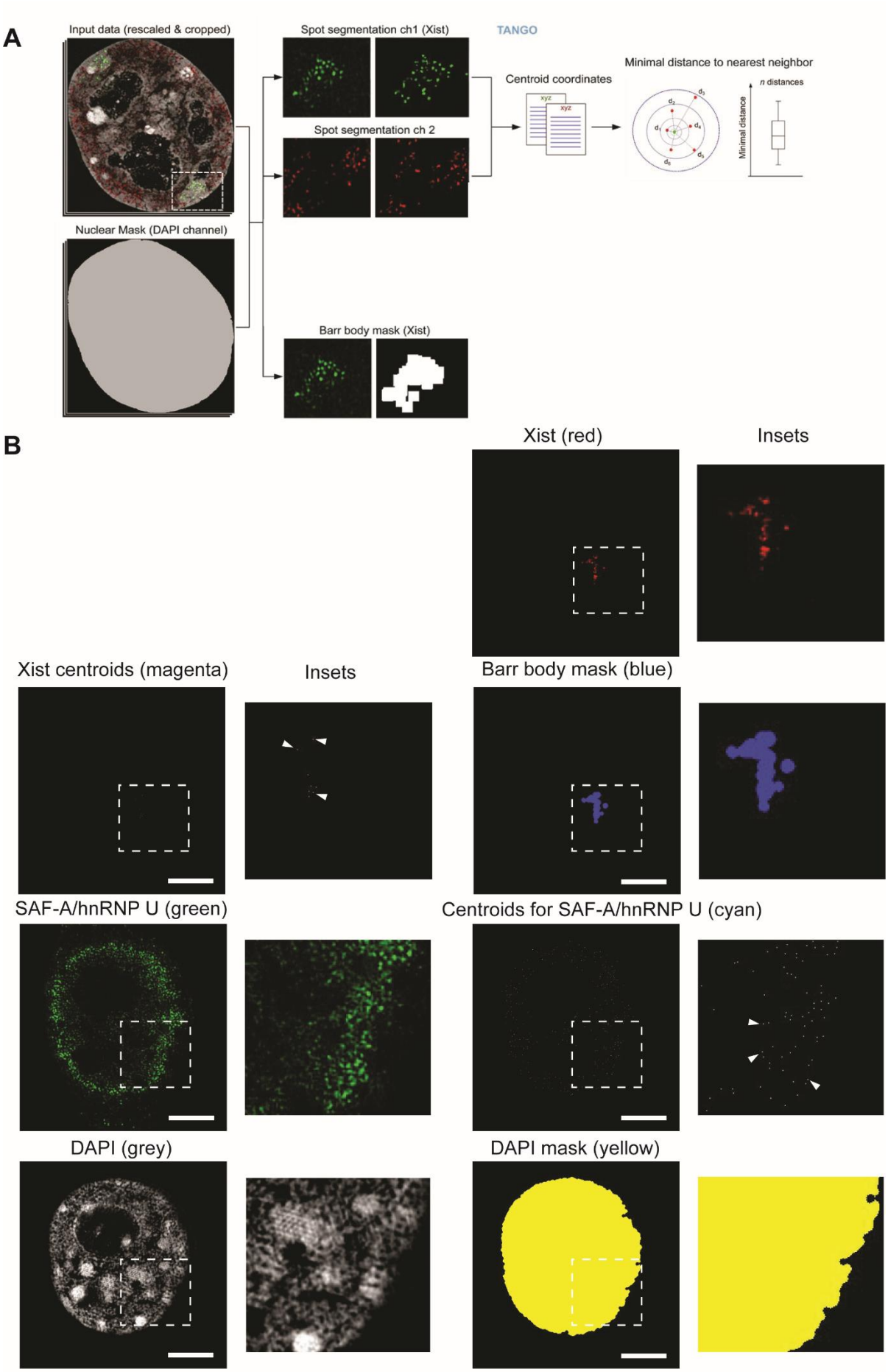

A
Input data (rescaled & cropped)
Spot segmentation ch1 (Xist)
TANGO
Spot segmentation ch 2
Centroid coordinates
xyz
xyz
Minimal distance to nearest neighbor
n distances
Minimal distance
Nuclear Mask (DAPI channel)
Barr body mask (Xist)
B
Xist (red)
Insets
Xist centroids (magenta)
Insets
Barr body mask (blue)
SAF-A/hnRNP U (green)
Centroids for SAF-A/hnRNP U (cyan)
DAPI (grey)
DAPI mask (yellow)

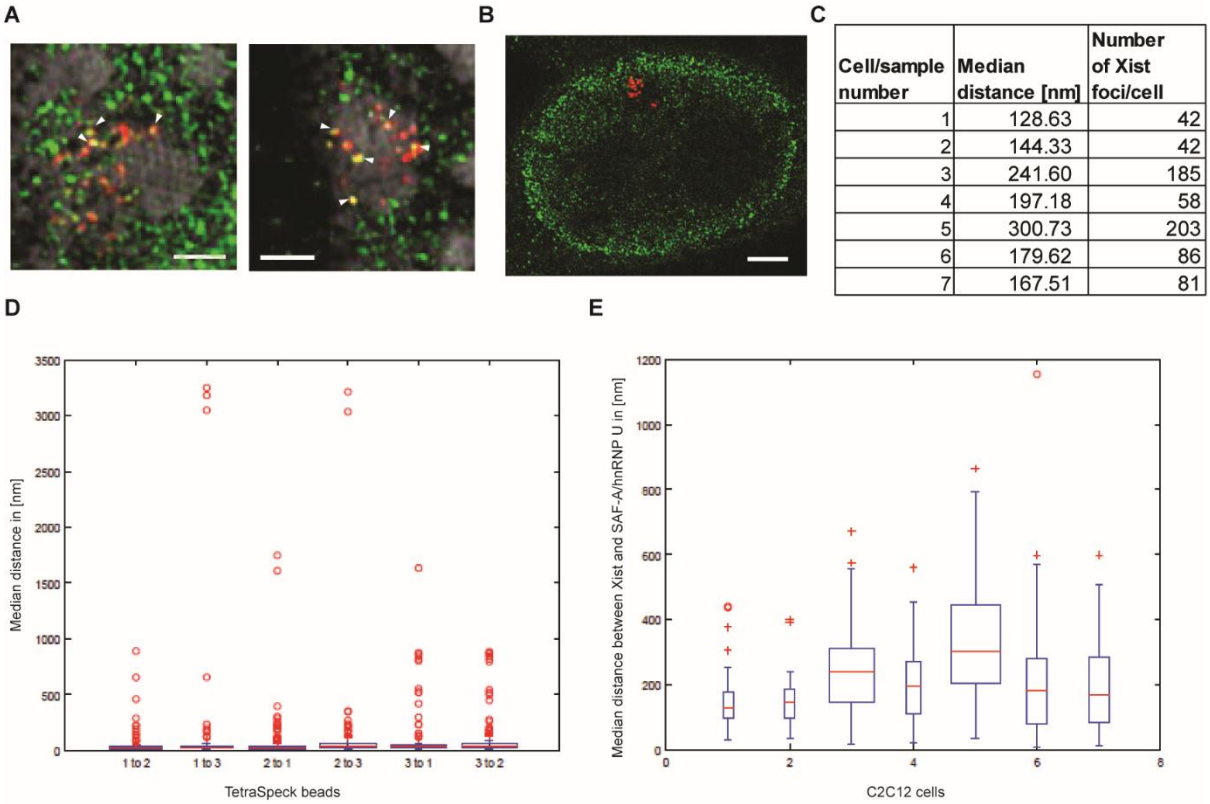
Figure 6.4 – 3D-SIM analysis workflow and example images.

(A) Schematic workflow for the quantitative analysis of 3D-SIM data showing a representative image of a C2C12 cell nucleus co-stained for Xist RNA (green) and murine SAF-A/hnRNP U (red), counterstained with DAPI (grey). Data were evaluated using the Fiji plugin TANGO (Ollion, Cochennec et al. 2013). Insets show 3D-SIM images (left) and corresponding results after segmentation/processing (right). Centroid coordinates were used for nearest neighbor analysis (adapted from (Smeets, Markaki et al. 2014)) (B) Sample images of a C2C12 cell stained with anti-SAF-A/hnRNP (green) and Xist probe (red). DNA/chromatin was counterstained with DAPI. Enlarged Xi regions of the cell are shown in the insets. Both Barr body (blue) and DAPI (blue) masks are shown as well. Centroids are marked by arrowhead (▼) for better visibility (Scale bars: 5 µm).

An initial visual examination of the images showed that there was some close association between Xist and SAF-A/hnRNP U at the Barr bodies of selected cells (Figure 6.5A). However, this observation was not entirely consistent across the whole population. The image in Figure 6.5B shows an example of a cell with no apparent overlap of Xist RNA and SAF-A/hnRNP U signals. This is owed to the obviously suboptimal SAF-A/hnRNP U staining pattern that in the present samples is characterized by a gradient from the nuclear periphery to the interior, and a relatively large amount of high-frequency artifacts. Particularly if the Xi extended into the nuclear interior, there were only very few, potentially unspecific, SAF-A/hnRNP U signals detected. This may be due to the fact that, in this set of experiments, a batch of antibody preparation different from the one that had been utilized in the Smeets *et al.* (2010) publication was used (Smeets, Markaki et al. 2014). In future experiments this issue can be easily addressed by titrating the amount of antibody used in the immuno-FISH preparations (personal communication with Dr. Lothar Schermelleh).

With respect to the present data, this technical artifact is reflected in the median distances derived by nearest neighbor analysis (Figure 6.5C and E). Here, they range from 128.63 nm to 300.73 nm. These numbers deviate from the published median distances of approximately

120 nm that were measured between the two macromolecules in the same cell line (Smeets, Markaki et al. 2014). Furthermore, the number of Xist domains per Xi determined here also does not match the published figures. Whereas Smeets *et al.* had observed 95 ±25 domains per Xi, the number of domains counted in this study varied a lot and was in the range of 42 to 203 domains (Figure 6.5C). Notably, the TetraSpeck bead measurements that are between 20 nm and 30 nm, presented in Figure 6.5D may be used as an indicator for the system error in the accuracy of spot distance measurements. Importantly, only the median values should be considered in Figure 6.5D as there are some artifactual results from the alignment procedure that creates areas at the rim with beads that have only one or two colors (personal communication with Dr. Lothar Schermelleh).



Cell/sample number	Median distance [nm]	Number of Xist foci/cell
1	128.63	42
2	144.33	42
3	241.60	185
4	197.18	58
5	300.73	203
6	179.62	86
7	167.51	81

Figure 6.5 – Spatial overlap of Xist RNA and endogenous SAF-A/hnRNP U at C2C12 Barr bodies.

(A) Two exemplary C2C12 Barr bodies with immuno-stained endogenous SAF-A/hnRNP U (green) and probe-labelled Xist RNA. Nuclei were counterstained with DAPI (grey). Arrowheads (▼) point at locations with a clear overlap of Xist RNA and SAF-A/hnRNP U signals (yellow) (Scale bars: 1 µm). (B) C2C12 whole nucleus stained with anti-SAF-A/hnRNP U (green) and Xist RNA probe (red) (Scale bar: 5 µm). (C) Seven selected images of C2C12 cells were subjected to nearest neighbor analyses. The table displays median distances between Xist RNA and SAF-A/hnRNP U foci and number of foci detected in a given cell. (D) Median distances between three different colors in TetraSpeck bead images. (E) Graphic representation of median distances between Xist RNA reference and SAF-A/hnRNP U foci in seven sample cells.

6.3 Summary and discussion

6.3.1 SAF-A/hnRNP U associate with Xist RNA

Even though the published results could not be replicated, the present data (Figure 6.5A) show that SAF-A/hnRNP U closely associates with Xist RNA (Smeets, Markaki et al. 2014) This does not mean that the two must interact directly, but it implies that their functionalities are linked. The question of how they are linked remains to be answered. Several scenarios can be envisioned. 1) There may be a direct interaction between SAF-A/hnRNP U and Xist RNA as suggested previously by Hasegawa *et al.* (2010) (Hasegawa, Brockdorff et al. 2010). 2) There may be an indirect protein-mediated interaction between SAF-A/hnRNP U and Xist RNA, and the two are part of a larger macromolecular complex. 3) There may be no direct interaction between the protein and the lncRNA. Instead, SAF-A/hnRNP U may function as a scaffolding protein only, and assists chromatin compaction and silencing at the Xi. By forming a reticular network around compacted chromatin it may physically entrap Xist RNA and keep it in close spatial proximity to the Xi. In light of the preliminary CLIP data presented in Chapter 3, the latter hypothesis appears to be most probable. Since no Xist RNA was found in the pull-down, it seems more likely that SAF-A/hnRNP U functions as a molecular mesh that retains Xist localization.

To substantiate initial evidence on SAF-A/hnRNP U's functional role further experiments need to be conducted. These may include 3D-SIM experiments with SAF-A/hnRNP U deletion variants, e.g. the protein without one of its functional domains. Looking at the localization patterns of these at high optical resolution may enhance our understanding of SAF-A/hnRNP U's roles in XCI and chromatin structure organization.

7. General discussion - A role for SAF-A/hnRNP U in mouse XCI?

7.1 SAF-A/hnRNP U nuclear localization

It had been shown previously that SAF-A/hnRNP U is enriched at and stably associates with the Xi (Helbig and Fackelmayer 2003, Fackelmayer 2005, Pullirsch, Hartel et al. 2010). Here, these findings were confirmed by studies on the *in vivo* localization of hrGFP-tagged murine SAF-A/hnRNP U in female somatic cell lines (Chapters 4). Furthermore, the protein appears to occur in two distinctive forms: one that predominantly locates to the Xi and another one that can be found almost everywhere in the nucleoplasm. As yet, the features differentiating the two protein pools have not been identified. They do seem to have an effect on SAF-A/hnRNP U recognition by antibodies. The antibody raised for the present study failed to detect an apparent enrichment of the scaffolding protein at the Xi; even in *in situ* nuclear matrix preparations (Chapter 2). With the exception of an autoimmune serum described by Pullirsch *et al*., only (hr)GFP-tagged SAF-A/hnRNP U could be visualized at the Xi as shown in this and previous works (Chapter 4) (Helbig and Fackelmayer 2003, Hasegawa, Brockdorff et al. 2010, Pullirsch, Hartel et al. 2010).

Amongst the possible explanations for this phenomenon are PTMs. These could induce changes in polypeptide conformation and lead to epitope exclusion. In this study, a proteomic approach was taken to identify putative modifications on SAF-A/hnRNP U (Chapter 2). Unexpectedly, already published PTMs like serine phosphorylation were not identified here (Britton, Froment et al. 2009). This may have been due to cell type- and species-specific differences or variable experimental conditions. Instead, it was found that SAF-A/hnRNP U is an extensively ubiquitinated protein (Chapter 2), an observation that is in line with its rapid turn-over during the cell cycle (Chapter 5). Another possible explanation would be the presence of multiple protein isoforms (Chapter 2). Indeed, it was demonstrated that soluble SAF-A/hnRNP U divides into a higher and a lower molecular weight fraction (Chapter 2). Thirdly, epitope exclusion may be caused by the tight packaging of Xi-bound

SAF-A/hnRNP U protein. This may be due to multimerization of SAF-A/hnRNP U at the Xi (Fackelmayer, Dahm et al. 1994). The possibility for multimerization had been underpinned by the published observation that the protein forms filamentous structures that outline the interchromatin space in the nucleus (Smeets, Markaki et al. 2014). However, it was also shown that filament formation is not exclusive to the Xi-territory, and occurs everywhere in the nucleus.

Considering the data presented here, it seems less likely that PTMs play a predominant role in the mouse model, and also filament formation does not appear to be specific to Xi-bound SAF-A/hnRNP U. Therefore, it can be suggested that different isoforms or protein multimerization impede antibody recognition.

7.2 SAF-A/hnRNP U dynamics

This and a previous study have investigated the turnover of interphase SAF-A/hnRNP U in the nuclear and the Xi-associated pool in mouse and human cells, respectively (Chapter 4 and Fackelmayer 2005). Both studies have come to the conclusion that SAF-A/hnRNP U stably associates with the Xi, with no conceivable dissociation within the imaging time. Moreover, it appears that the nucleoplasmic SAF-A/hnRNP U pool cannot readily be transformed into a Xi-binding form.

Its very stable interphase association hints at a structural role for SAF-A/hnRNP U in XCI, possibly by creating a distinctive, silent compartment and physically entrapping Xist RNA.

To further the insights into this process, SAF-A/hnRNP U dynamics were studied throughout the cell cycle (Chapter 5). It was shown that SAF-A/hnRNP U dissociates from the Xi prior to Xist RNA. This observation clearly argues against the direct SAF-A/hnRNP U-Xist interaction that had been suggested in earlier works (Hasegawa, Brockdorff et al. 2010).

Furthermore, imaging of the Xi-enriched proteins SAF-A/hnRNP U and PCGF3 in live-cells showed that they behave similarly during mitosis (Chapter 5). This observation may suggest a similar role in XCI for the two. Even though PCGF3 was not found to associate with SAF-A/hnRNP U in Xist-induced ESCs, it may be that they form a complex in the somatic FSPE cell line.

7.3 SAF-A/hnRNP U RNA binding

The SAF-A/hnRNP U RNA binding behavior was investigated using CLIP-Seq (Chapter 3). The preliminary results presented here suggest that the protein is not a direct Xist RNA binder and contradict the published literature (Hasegawa, Brockdorff et al. 2010). However, it cannot be excluded that the negative CLIP-Seq outcome has originated from flaws in the experimental setup. Amongst others, it had been shown by microscopy that the antibody used for CLIP did not highlight the Xi enrichment of SAF-A/hnRNP U (Chapter 2). Taken together with the observation that the pan-nuclear and the Xi-bound pool of SAF-A/hnRNP U are distinct, this may indicate that the in house antibody prepared for this study is an inadequate reagent for such experiments. This was also underpinned by 3D-SIM imaging that showed an apparent depletion of antibody-stained protein in the Xi territory. In contrast the hrGFP-tagged variant was clearly enriched (Chapter 4). Therefore, in the future full CLIP-Seq experiments with all appropriate controls and SAF-A/hnRNP U:hrGFP need to be performed.

The aforementioned 3D-SIM experiments and data in the published literature do show that SAF-A/hnRNP U and Xist RNA associate in mouse cells implying a functional connection between the two (Chapter 6) (Smeets, Markaki et al. 2014). Smeets *et al.* (2014) showed that the mean minimal distance between Xist RNA and SAF-A/hnRNP U is about 120 nm, which is lower than the minimal distances for Xist RNA and the PRC2/1 components Ezh2 (140 nm), Suz12 (158 nm) and Ring1b (140 nm) determined by Cerase *et al.* (2014) (Cerase, Smeets et al. 2014, Smeets, Markaki et al. 2014). Because of the rather large

separation between PRC2/1 components and Xist RNA, Cerase *et al.* suggest two possible explanations for PRC2/1-Xi associations: On the one hand, there may be an RNA binding adapter protein that binds the silencing complexes and establishes an indirect interaction between PRC2/1 and the lncRNA. On the other hand, Xist-induced changes in chromatin structure may account for PRC2/1 recruitment. SAF-A/hnRNP U may be involved in either. Even though the preliminary CLIP data presented here do not favor a direct interaction between SAF-A/hnRNP U and Xist RNA, there is still the possibility that such an interaction exists in the living cell. However, more evidence needs to be provided to substantiate this idea (Hasegawa, Brockdorff et al. 2010). As for the second hypothesis, undoubtedly SAF-A/hnRNP U plays some role in chromatin organization and therefore may be one of the factors that create a permissive environment for PcG protein recruitment.

Different models for SAF-A/hnRNP U molecular associations in conjunction with Xist RNA can be built upon the experimental evidence collected to date (Figure 7.1). Firstly, SAF-A/hnRNP U may bind Xist RNA indirectly through a bridging adapter protein, for example a small scaffold protein (Figure 7.1A). It may be that only a specific SAF-A/hnRNP U isoform can be bound by the adapter. This model is the one favored by the CLIP results (Chapter 3). However, it may well be that SAF-A/hnRNP U directly binds to Xist RNA and is then recruited to the Xi by another SAF-A/hnRNP U-binding protein to the Xi (Figure 7.1B). These two models assume that SAF-A/hnRNP U does not necessarily bind to RNA and DNA simultaneously. Thirdly, SAF-/hnRNP U may be part of an Xi-bound protein complex and bind to neither Xist RNA nor Xi chromatin directly (Figure 7.1C). Naturally, this does not exclude the possibility that interactions within this complex are RNA-mediated. Lastly, there remains the possibility that SAF-A/hnRNP U concurrently interacts with Xist RNA and chromatin (Figure 7.1D), or is not involved in Xist binding and recruitment at all.

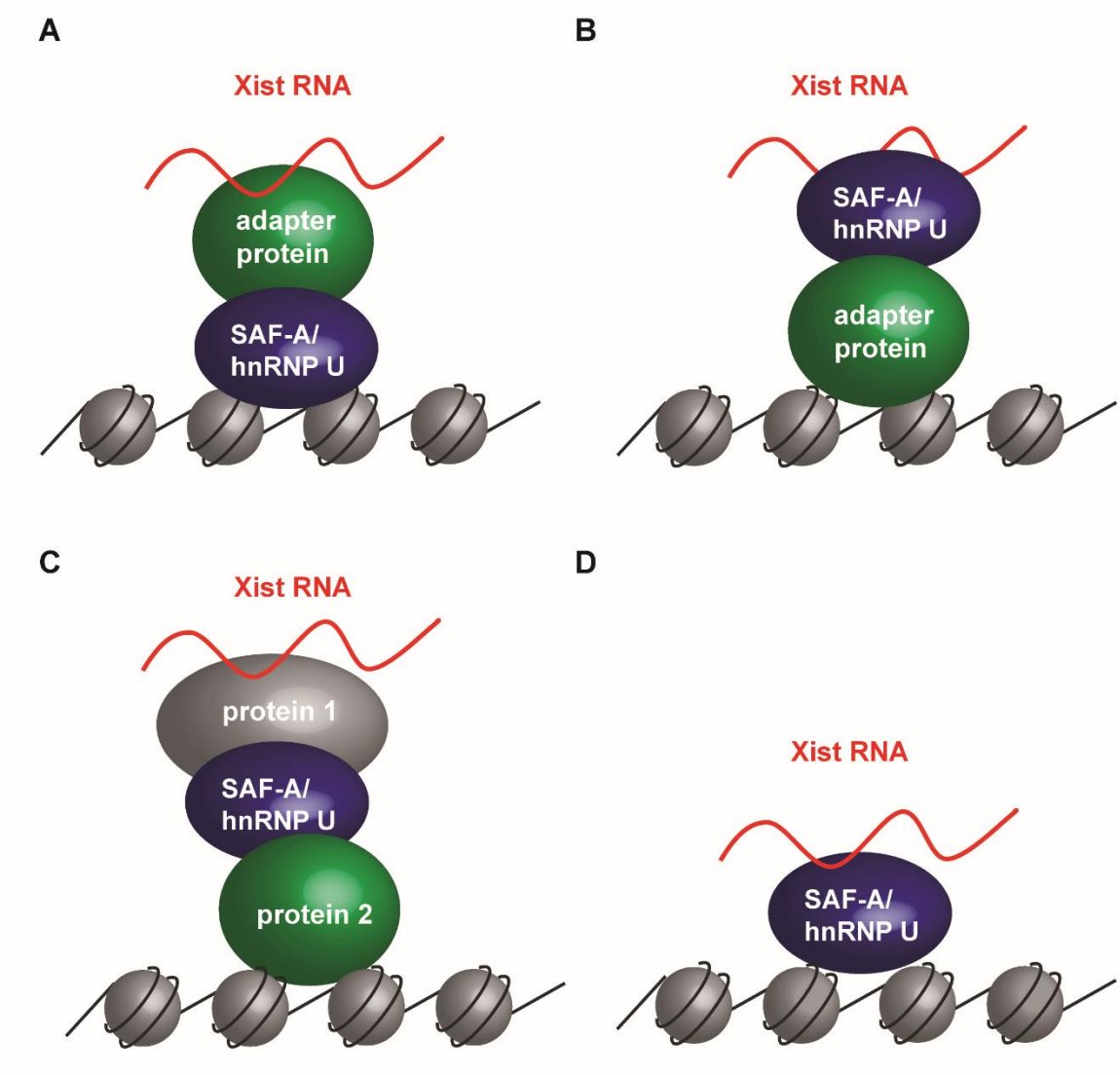


Figure 7.1 – Models for SAF-A/hnRNP U molecular associations.

(A) SAF-A/hnRNP U attaches to chromatin sites and recruits an Xist-bound adapter protein (complex). (B) SAF-A/hnRNP U itself binds Xist RNA and the complex locates to the Xi through a specific adapter protein (complex). (C) SAF-A/hnRNP U is an adapter protein sandwiched between an Xist binding and a chromatin binding factor. (D) SAF-A/hnRNP U directly binds to Xist RNA and tethers it to the Xi chromatin.

In summary, the results of this thesis support the idea that SAF-A/hnRNP U acts as a structural component that plays an important role in nuclear organization. It seems to prevent free, non-directional RNA diffusion throughout the nucleus, and its ability to

multimerize may support this function (Fackelmayer, Dahm et al. 1994). The prominent enrichment of SAF-A/hnRNP U at the Xi pointed to a role for the protein as an Xist associating factor. This thesis work proposes that SAF-A/hnRNP U does not directly interact with the lncRNA, but instead forms a barrier that prevents unrestricted RNA movement (Figure 7.2). This function for SAF-A/hnRNP U in the regulation of nuclear RNA localization has also been reported in a recent study that investigated chromosome associations of C_0T-1 repeat RNAs (Hall, Carone et al. 2014). Likewise the SAF-A/hnRNP U barrier may also prevent entry of transcription factor to the Xi territory, and may be involved in promoting silencing mark (e.g. H3K27me3) stability by prohibiting access for their removal factors. In the future this hypothesis needs to be scrutinized experimentally to further our understanding of XCI.

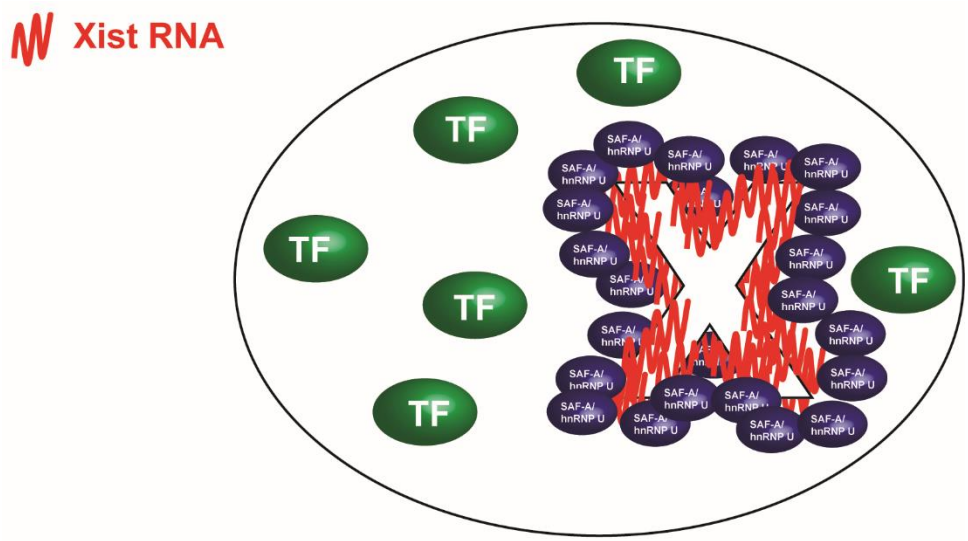


Figure 7.2 – SAF-A/hnRNP U confines Xist RNA to the Xi territory.

SAF-A/hnRNP U may be a structural protein constraining Xist RNA diffusion within the nucleus by creating a secluded nuclear compartment. Transcription factors (TF) may not access this compartment.

7.4 Additional considerations

7.4.1 Genetic manipulation of SAF-A/hnRNP U or Xist RNA expression

One of the shortcomings of the present study is the lack of SAF-A/hnRNP U knockdown (KD), knockout (KO) or partial deletion experiments.

SAF-A/hnRNP U KD experiments were performed in the Nakagawa laboratory (Hasegawa, Brockdorff et al. 2010), where researchers achieved near-complete siRNA-mediated KD of the target protein in Neuro2a cells, and showed that SAF-A/hnRNP U KD interfered with focal Xist localization. Rescue experiments utilizing FLAG-tagged SAF-A/hnRNP U showed a re-recruitment of Xist RNA to the Xi in those cells. On the contrary, when KD cells were supplemented with SAF-A/hnRNP U deletion constructs where either the DNA or the RNA binding domain was missing (ΔSAF or ΔRGG), Xist RNA did not re-locate to the Xi (Hasegawa, Brockdorff et al. 2010). Additional partial deletion studies with SAF-A/hnRNP U:hrGFP were conducted by Pullirsch *et al.* (2010) who examined SAF-A/hnRNP U localization upon deletion of the DNA binding SAF box in HEK 293T cells (Pullirsch, Hartel et al. 2010). Their experiments showed that SAF-A/hnRNP U:hrGFP without the SAF box was no longer enriched at the Xi.

To date, SAF-A/hnRNP U KO cells have not been utilized to study the relationship between SAF-A/hnRNP U and Xist RNA. Initial attempts to generate hypomorphic SAF-A/hnRNP U mutant mice using an U3Neo gene trap system failed, due to early embryonic lethality (Roshon and Ruley 2005). Instead, a mouse cell line with a hypomorphic SAF-A/hnRNP U variant was generated (Roshon and Ruley 2005). And only recently, the group around Tom Maniatis (2015) published evidence showing that his laboratory had generated heart-specific, conditional SAF-A/hnRNP U KO mice (Ye, Beetz et al. 2015).

A conditional SAF-A/hnRNP U KO cell line would have been a useful tool for this thesis work. With the advent of the clustered regularly interspaced short palindromic repeat (CRISPR)-based techniques in recent years (Taylor and Woodcock 2015), generating such

a line should be a lot more feasible than in the past. A conditional KO model would have been useful in SAF-A/hnRNP U:hrGFP complementation experiments, and would also be a good control for various experiments presented here, i.e. CLIP and 3D-SIM imaging.

But not only SAF-A/hnRNP U KO models would have been suitable tools for this thesis work, also the usage of conditional Xist KO systems may have added significance to experimental outcomes. Xist KO studies have been conducted by Pullirsch *et al.* (2010) (Pullirsch, Hartel et al. 2010), who prepared NIH 3T3 cells homozygous for a conditional *Xist* allele bearing an inducible *Cre*-transgene that had been stably transfected with a SAF-A/hnRNP U:hrGFP fusion construct. Xi-enrichment of SAF-A/hnRNP U was observed prior to Xist deletion. However, upon induction of the *Cre*-transgene focal enrichment of SAF-A/hnRNP U:hrGFP at the Xi was lost (Pullirsch, Hartel et al. 2010). Complementary experiments showed that in a transgenic ESC line Xist-induction combined with differentiation lead to focal enrichment of SAF-A/hnRNP U:hrGFP at the site of *Xist*-transgene expression in a proportion of cells (Pullirsch, Hartel et al. 2010). Moreover, loss of *Xist* expression during differentiation was concomitant with the depletion SAF-A/hnRNP U:hrGFP from the sites of focal enrichment. Altogether these findings suggest that SAF-A/hnRNP U recruitment is dependent on Xist RNA expression and focal enrichment.
Xist KO experiment would have complemented experiments under Xist induced conditions here. Furthermore, Xist KO cells could have been used to verify Pullirsch *et al.*'s findings in the cell lines used for this thesis work.

7.4.2 SAF-A/hnRNP U chromatin immunoprecipitation

SAF-A/hnRNP U had been identified as component of the nuclear scaffold (Kiledjian and Dreyfuss 1992, Romig, Fackelmayer et al. 1992). One typical characteristic of the protein is its ability to bind to S/MARS, which are DNA elements of the eukaryotic genome that are believed attach chromatin fibres to the nuclear matrix (Gohring and Fackelmayer 1997, Liebich, Bode et al. 2002).

In light of SAF-A/hnRNP U's specific enrichment at the Xi it would have been interesting to know whether the Xi-specific binding was evident in chromatin immunoprecipitation (ChIP). Such an experiment would require a female cell line bearing two distinguishable X chromosomes, ideally combined with skewed X chromosome inactivation, i.e. inactivation of one of the chromosomes is strongly favoured over the other. Possible model systems would be differentiated PGK12.1 mouse embryonic stem cells, whose X chromosomes can be distinguished easily by single nucleotide polymorphism (SNP) analysis, or mouse cells that bear a T(X;16)16H translocation also known as Searle's translocation that leads to non-random XCI (Norris, Patel et al. 1994, Takagi, Sugimoto et al. 2002, Lin, Gupta et al. 2007). After initial optimization of the ChIP conditions, a ChIP-Seq experiment with subsequent bioinformatics analysis may reveal any differences in SAF-A/hnRNP U binding to DNA regions of the Xi and Xa, respectively.

Initial attempts to optimize the ChIP conditions using the in house SAF-A/hnRNP U antibody and PGK12.1 cells have been made in the course of this thesis work (data not shown). In preliminary experiments primers were designed for ChIP-qPCR using an S/MAR prediction algorithm (http://genomecluster.secs.oakland.edu/marwiz). The original primer design was complicated by the fact that typically S/MARS are found in repeat-rich regions of the genome and the low predictive power of S/MAR prediction algorithms (Evans, Ott et al. 2007).

Once initial primer tests had been completed, this part of the project had not been pursued any further.

In the future experiments addressing the question of whether differential binding of SAF-A/hnRNP U to the Xa and Xi can be shown by ChIP-Seq would make major contributions to elucidating whether there is a functional role for SAF-A/hnRNP U in the XCI process. Ideally this set of experiments would be conducted with an antibody that highlights the enrichment of SAF-A/hnRNP U at the Xi in IF imaging. Furthermore, it is important that all appropriate controls are included. This means that apart from the standard IgG antibody control, SAF-A/hnRNP U DNA binding behavior in differentiated and undifferentiated PGK12.1 cells

should be compared, and that for the female T(X;16)16H cell line a male control line should be included in the experimental setup.

8. Materials and Methods

Abbreviations

aa	amino acid(s)
ATP	adenosine triphosphate
BSA	bovine serum albumin
CBB	Coomassie Brilliant Blue G-250
cDNA	complementary DNA
CLIP	*in vivo* UV crosslinking and immunoprecipitation
DAPI	4',6-diamidino-2-phenylindole
DMSO	dimethyl sulfoxide
dox	doxycycline
DTT	dithioreitol
ECL	electrochemiluminescence
EDTA	ethylenediaminetetraacetic acid
EMCCD	electron multiplying coupled charge device
ESC	embryonic stem cell
ESI-QUAD-TOF	electrospray ionization-quadrupole-time-of-flight
FCS	fetal calf serum
FDR	false discovery rate
FGF4	fibroblast growth factor 4
FSG	fish skin gelatin

FSPE	female *mus spretus* ear embryonic fibroblasts
g (-force)	gravitational force
GO	gene ontology
h	hour(s)
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HF	high fidelity
His	Histidine
HITS	high-throughput sequencing
HRP	horseradish peroxidase
HRP	horseradish peroxidase
IF	immunofluorescence\
IPTG	isopropyl-β-D-thiogalactosylpyranoside
LB	lysogeny broth
LIC	ligation-independent cloning
LIF	leukemia inhibitory factor
min	minute(s)
MW	molecular weight
NaTT	sodium tetrathionate
ne	nuclear extract(s)
NP-40	Nonidet P-40
N-terminus	amino terminus
OD_{600}	optical density at wavelength of 600 nm
PAGE	polyacrylamide gel electrophoresis

PBS	phosphate buffered saline
PCR	polymerase chain reaction
PEF	primary embryonic fibroblasts
PIPES	piperazine-N,N′-bis(2-ethanesulfonic acid)
plgem	power law global error model
PVDF	polyvinyl diflouride
qPCR	quantitative real-time PCR
ROI	region of interest
rpm	revolutions per minute
s	second(s)
SC	spectral count
SDS	sodium dodecyl sulfate
SINQ	spectral index quantification
SSC	saline-sodium citrate
TBST	Tris-buffered saline and Tween 20
T_m	melting temperature
TSC	trophoblast stem cell
v/v	volume to volume
v/w	volume to weight
wce	whole cell extract(s)

8.1 Molecular cloning

All Polymerase chain reaction (PCR) primers were designed using VectorNTI (InforMax) and/or SeqBuilder (DNA Star) and publically available genome databases. The conventional sequencing was done with Source Biosciences. Subsequent sequence alignments were done with BioEdit (Tom Hall).

8.1.1 PCR reaction setup

Inserts for cloning were amplified using either Phusion High-Fidelity (HF) DNA Polymerase (NEB) or Advantage HD DNA Polymerase Mix (Clontech). The reactions were set up according to the supplier's instructions and run on a DNA Engine Dyad Thermal Cycler (Bio-Rad). PCR cycling conditions were adjusted to the primers' (T_m) and the length of the expected PCR product. Colony PCRs were conducted with Taq DNA Polymerase (Invitrogen) following the supplier's instructions.

8.1.2 LIC cloning of N-terminal murine SAF-A/hnRNP U

The N-terminal (amino-terminal) part (amino acids [aa] 1 to 260) of the murine SAF-A/hnRNP U was cloned from a pcDNA3 vector containing SAF-A/hnRNP U into a pNIC28-Bsa4 construct using ligation-independent cloning (LIC) (Aslanidis and de Jong 1990). 5'-TACTTCCAATCCATGAGTTCTTCGCCTGTT-3' and 5'-TATCCACCTTTACTGATCGAAGTGTTCGTC-3' were used for priming in forward and reverse direction, respectively. Selected clones were screened by colony PCR and restriction digests. Positive clones were verified by sequencing.

8.1.3 TOPO cloning

The TOPO cloning was carried out using the StrataClone PCR Cloning Kit (Agilent Technologies) following the supplier's instructions. White colonies were selected for colony PCR with T3 (5'-ATTAACCCTCACTAAAGGGA-3') and T7 (5'-TAATACGACTCACTATAGGG-3') primers. Colonies of interest were verified by sequencing with T3 primers.

8.2 Protein expression and purification

The expression constructs were transformed into chemically competent BL21DE3 cells (in house preparations). A medium-sized single colony was picked and used to grow a pre-culture in lysogeny broth (LB) under appropriate antibiotic selection. The pre-culture was used to inoculate the main culture that was grown up to an OD_{600} of 0.6 before induction with 1 mM isopropyl-β-D-thiogalactosylpyranoside (IPTG). Prior to harvesting the overexpression culture was either grown for 4 h at 30 °C or kept shaking overnight at 20 °C. Eventually, cells were collected by centrifugation for 20 min at 4,000 g. After removing the supernatant the pellet was washed with phosphate buffered saline (PBS) and stored at -80 °C.

8.2.1 Purification of six Histidine-tagged proteins

All proteins cloned into the pNIC28-Bsa4 vector were fused to an N-terminal six Histidine-tag (His-tag), and could therefore be purified over a Ni^{2+}-NTA resin using the below protocol.

The protein pellets were re-suspended in His-lysis buffer (20 mM pH 8.0, 250 mM NaCl, 0.1 % (v/v) NP-40, and Mini ethylenediaminetetraacetic acid (EDTA)-free Protease Inhibitor Cocktail [Roche]). Then the cells were ruptured by sonication (40 % amplitude, 10 cycles 30 s on/off). Soluble and insoluble fractions were separated by centrifuging the crude lysate for 1 h at 20,000 rpm and the supernatant was applied to Ni^{2+}-chelated IMAC Sepharose Fast Flow resin (GE Healthcare). Resin and supernatant were left on a rotating wheel at 4 °C for at least 1 h. Following incubation the lysate and the beads were transferred into a PolyPrep Chromatography column (Bio-Rad) and washed with 20 - 40 column volumes of wash buffer (50 mM NaH_2PO_4 pH 8.0, 300 mM NaCl, and 20 mM imidazole). Finally, the protein was eluted with elution buffer (50 mM NaH_2PO_4 pH 8.0, 300 mM NaCl, and 250 mM imidazole). When appropriate fractions were pooled and dialyzed overnight against dialysis buffer (50 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid [HEPES] pH 7.9, 100 mM KCl. 10 % (v/v) glycerol, and 0.5 mM dithioreitol (DTT)) or PBS.

8.3 Antibody preparation and characterization

8.3.1 Antibody production

The polyclonal SAF-A/hnRNP U antibody was raised by injecting four times 150 µg His-tagged recombinantly expressed murine N-terminal SAF-A/hnRNP U (aa 1 to 260) protein over a period of four times 28 days into the rabbit. The immunization procedure had been conducted by the Scottish National Blood Transfusion Service PTU/BS (NHS). After receiving the last bleed the rabbit was exsanguinated and the spleen was removed and stored at -80 °C.

8.3.2 Antibody affinity purification

The serum derived from each of the four bleeds was purified by affinity purification against the N-terminal SAF-A/hnRNP U peptide (aa 1 to 260). Prior to the antibody purification 10 mg of the peptide were coupled to 500 µL Affi Gel-15 affinity support (Bio-Rad) by incubating the protein together with resin overnight at 4 °C. Next the Affi Gel-15 beads were blocked with 200 mM ethanolamine for 2 h at 4 °C. After transferring the resin to a PolyPrep Chromatography column (Bio-Rad) the beads were washed with 1 M NaCl and ice-cold PBS. 10 mL of crude rabbit serum were applied to the column. The affinity support and serum were incubated together for 3 h at room temperature. Afterwards the resin was washed with 500 mM NaCl followed by a PBS wash. The SAF-A/hnRNP U antibodies were eluted in several 500 µL fractions with 0.1 M glycine, pH 3.0. Immediately afterwards the eluate was neutralized by adding 50 µL of 1 M Tris, pH 8.0, to each of the fractions. Samples of the elution fractions were run on a SDS-PAGE gel, pooled and dialyzed into buffer containing 50 mM NaCl and 10 % (v/v) glycerol. Also, some of the fractions were further concentrated using Vivaspin 20 columns (Sartorius stedim). Finally, the antibody was aliquoted and stored at -80 °C.

8.3.3 Antibody-antigen competition assay

A total of 2 µg anti-SAF-A/hnRNP U antibody was used per immunofluorescence (IF) on FSPE cells grown on coverslips. Prior to applying the antibody to the fixed and permeabilised cells the SAF-A/hnRNP U antigen (amino acids 1 to 260) was added at different concentrations and left for incubation (5 min, room temperature). The following steps were conducted as described in the section “Indirect immunofluorescence”. Eventually, all images in the antibody-antigen competition assay were taken with an Axio Observer Z1 microscope and processed with the AxioVision Rel. 4.8 software (Zeiss) using the same exposure time for every image.

8.4 Antibodies

Table 8.1 displays the antibodies used in this study. Shown are also the antibody supplier names, catalogue number, type (mono- or polyclonal), stock concentrations and working dilutions. Secondary antibodies have been previously characterized in the Brockdorff lab, i.e. in IF experiments with secondary antibodies only.

Table 8.1 – Antibodies.

Antibodies with supplier names, catalogue number, type (mono-or polyclonal), stock concentrations and working dilution. The numbers 1 and 2 indicate whether the reagent was a primary or secondary antibody, respectively.

	Antibody	Supplier	Catalogue Number	Poly-clonal	Mono-clonal	Stock concentration	Working Dilution
Primary Antibodies	rabbit anti-SAF-A/hnRNP U (**1**)	n/a	n/a	x		0.5 mg/mL	1:10,000, IF 1:100 - 1:200, IP 0.5 - 2 ug
	mouse anti-H3K27me3 (**1**)	Millipore	04-801		x	n/a	IF 1:100
	mouse anti-H3K27me3 (**1**)	Diagenode	#MAb-014-050		x	1.5 mg/mL	IF 1:100
	mouse IgG1 anti-His (**1**)	Santa Cruz	sc-8036		x	n/a	WB 1:200
	rabbit anti-mouse IgG (**1**)	Sigma	M7023	x		2.8 mg/mL	IP 5.6 ug
	rabbit anti-hrGFP (**1**)	Agligent Technologies	240141	x		n/a	1:5,000, IF 1:200
	mouse anti-FLAG M2 (**1**)	Sigma	F1804		x	1.0 mg/mL	WB 1:1,000, IF1:100 - 1:200
Secondary Antibodies	HRP donkey anti-rabbit IgG (**2**)	GE Healthcare	NA934V	x		0.21 mg/mL	1:20,000
	HRP donkey anti-mouse IgG (**2**)	GE Healthcare	NA931V	x		0.21 mg/mL	WB 1:10,000 - 1:20,000
	Alexa Fluor 568 goat anti-mouse IgG (**2**)	Molecular Probes	A11031		x	2.0 mg/mL	IF 1:400
	Alexa Fluor 568 goat anti-rabbit IgG (**2**)	Molecular Probes	A11011		x	2.0 mg/mL	IF 1:400
	Alexa Fluor 594 goat anti-rabbit IgG (**2**)	Molecular Probes	A21207	x		2.0 mg/mL	3D-SIM 1:1,000
	Alexa Fluor 488 goat anti-mouse IgG (**2**)	Molecular Probes	A11029		x	2.0 mg/mL	IF 1:400
	Alexa Fluor 488 goat anti-rabbit IgG (**2**)	Molecular Probes	A11008		x	2.0 mg/mL	IF 1:400, 3D-SIM 1:1,000
	IR Dye 800CW conjugated goat anti-rabbit IgG (**2**)	Li-COR	926-32211	x		1.0 mg/mL	WB 1:15,000
	IR Dye 680RD conjugated goat anti-rabbit IgG (**2**)	Li-COR	925-32213	x		1.0 mg/mL	WB 1:15,000

8.5 Protein biochemistry

8.5.1 Whole cell extract preparation

Whole cell extracts (wce) were prepared suspending cells in RIPA buffer (1% (v/v) NP-40, 150 mM NaCl, 10 % (v/v) Glycerol, 50 mM Tris pH 8.0, 1 mM DTT, Mini EDTA-free Protease Inhibitor Cocktail [Roche]) and leaving the cells for lysing on a rotator at 4 °C for a minimum of 30 min. Subsequently, the lysate was spun 15 min at 13,000 rpm, 4 °C and the supernatant was taken as wce.

8.5.2 Nuclear extract preparation

Cells were harvested by trypsinization, washed with PBS and subsequently re-suspended in five times pellet volume buffer H (20 mM HEPES pH 7.9, 1.5 mM $MgCl_2$, 10 mM KCl, 1 mM DTT, EDTA-free Protease Inhibitor Cocktail [Roche]). The cell suspension was passed five times through a 27G syringe needle and nuclei were spun down by centrifugation at 13,000 rpm, 5 min). The pellet was re-suspended in five times pellet volume buffer D (20 mM HEPES pH 7.9, 1.5 mM $MgCl_2$, 0.2 mM EDTA, 0.42 M KCl, 20 % (v/v) glycerol, 1 mM DTT, EDTA-free Protease Inhibitor Cocktail [Roche]). Nuclear proteins were extracted by vigorously shaking the nuclei with a Fast-Prep24 (MP Biomedicals) for 60 sec. The nuclear protein fraction was cleared by centrifugation at 13,000 rpm, for 20 min at 4 °C. The total

protein concentration was estimated using a NanoDrop spectrophotometer (Thermo Scientific).

8.5.3 *In situ* nuclear matrix preparation

Prior to starting the nuclear matrix preparation FSPEs grown on coverslips were washed with cold PBS twice. Then the cells were overlaid with 100 µL matrix buffer 1 (10 mM piperazine-N,N′-bis(2-ethanesulfonic acid) [PIPES], pH 6.8, 100 mM NaCl, 300 mM Sucrose, 3 mM $MgCl_3$, 0.5 % (v/v) Triton X-100, 2 mM sodium tetrathionate [NaTT] freshly added) and incubated for 10 min at room temperature. This was followed by another 10 min incubation with matrix buffer 2 (10 mM PIPES, pH 6.8, 250 mM $(NH_4)_2SO_4$, 300 mM Sucrose, 3 mM $MgCl_3$, 0.5 % (v/v) Triton X-100) and a 20 min incubation with matrix buffer 3 (10 mM PIPES, pH 6.8, 50 mM NaCl, 300 mM Sucrose, 3 mM $MgCl_3$, 0.5 % (v/v) Triton X-100, 200 U/mL DNase I). Finally the cells were kept for 5 min in matrix buffer 4 (10 mM PIPES, pH 6.8, 50 mM NaCl, 25 mM $(NH_4)_2SO_4$, 300 mM Sucrose, 3 mM $MgCl_3$, 0.5 % (v/v) Triton X-100, 200 U/mL DNase I) and matrix buffer 5 (10 mM PIPES, pH 6.8, 1.7 mM NaCl, 300 mM Sucrose, 3 mM $MgCl_3$, 0.5 % (v/v) Triton X-100), respectively. When scaffold preparation were followed by RNA-FISH, all extraction buffers were supplemented with RNasin Plus RNase Inhibitor (40 U/µL, Promega) at 1 µL/mL. Eventually, cells were fixed with formaldehyde and either mounted directly onto microscopy slides or subjected to RNA-FISH or immunofluorescence staining. Images were taken with an Axio Observer.Z1 microscope and processed with the AxioVision Rel. 4.8 software (Zeiss) and Photoshop (Adobe).

8.5.4 SDS-PAGE

Proteins were denatured by boiling them for 5 to 10 min in Laemmli buffer (Laemmli 1970) and size-separated on SDS-PAGE gels. Dual Color Precision Plus Protein Marker (Bio-Rad) was used molecular weight marker. The resolved proteins were stained with Instant Blue (TripleRed) following the manufacturer’s instructions.

8.5.5 Western blotting

Unstained SDS-PAGE gels were soaked in western transfer buffer (20 % (v/v) methanol, 9.6 mM Tris, 7.8 mM glycine, 0.5 % (v/v) SDS). Similarly, polyvinyl diflouride (PVDF) membranes (GE Healthcare) were first washed with 100 % methanol, then washed with water. In some cases PVDF membranes (GE Healthcare) were substituted with nitrocellulose membranes (GE Healthcare). These did not a require pre-washing with 100 % methanol. Eventually, the blotting membranes and extra-thick blotting paper (Bio-Rad) were soaked in transfer buffer. Semi-dry transfer (Bio-Rad Transblot SD Semidry Transfer Cell) was conducted at 15 V for 15 to 30 minutes. Upon transfer the blotting membranes were blocked in 5 % (w/v) non-fat dried milk in TBST buffer (100 mM Tris-HCl pH8.0, 154 mM NaCl, 0.1 % (v/v) Tween-20). Primary antibodies were used at the dilutions indicated in Table 8.1 and incubated for 1 h at room temperature or at 4 °C overnight in TBST plus 0.5 % (w/v) non-fat milk. After three in-between washes species-specific HRP-coupled secondary antibodies (Suppl. Table 1) in 0.5 % (w/v) non-fat milk containing TBST were applied and incubated with the membranes for 30 min to 1 h.

Finally, signals were detected by adding electrochemiluminescence (ECL) reagent (GE healthcare) substrate and exposing X-ray films (GE Healthcare). Where applicable, band intensities were quantified with ImageJ/Fiji software and image brightness and contrast was adjusted using Photoshop (Adobe) (Abramoff 2004).

Western blot band intensity quantification was done either using the ImageJ software or Li-COR Image Studio. After background signal subtraction signal intensities were converted into relative percentages.

8.5.6 Immunoprecipitation

Nuclear extracts or whole cell extracts were diluted with BC150 (50 mM HEPES pH 7.9, 150 mM KCl, 10 % (v/v) glycerol, 0.5 mM EDTA, 0.5 mM DTT) and a fraction of the input was

saved. 5.6 µg rabbit polyclonal IgG (Sigma, M7023) and varying amounts of affinity purified SAF-A/hnRNP U antibody were added to the immunoprecipitation samples and incubated on a rotating wheel for 2 h or overnight at 4 °C. For pulling out the antibodies 20 µL of packed salmon sperm DNA-blocked protein A beads (Millipore) or magnetic protein A Dynabeads (Invitrogen) were added to the samples. Beads and immunoprecipitation samples were left rotating for 1 h at 4 °C. In the following the protein A resin was spun down (1,000 x g, 1 min, 4 °C) and washed three times with ice-cold BC300 (50 mM HEPES pH 7.9, 300 mM KCl, 10 % (v/v) glycerol, 0.5 mM EDTA, 0.5 mM DTT). Afterwards, the bound protein was eluted by boiling the beads in 30 µL Laemmli buffer. Following elution all immunoprecipitation samples were subjected to SDS-PAGE and Western blotting, both conducted as outlined in previous paragraphs (SDS-PAGE and Western blotting).

8.5.7 *In vivo* UV crosslinking and immunoprecipitation

The *in vivo* UV crosslinking and immunoprecipitation experiment (CLIP) was carried out following the protocol described by *Wang et al.* (Wang, Tollervey et al. 2009). Briefly, Xist-inducible 3E cells were grown on 10 cm tissue culture dishes and induced with doxycycline (dox) for 48 h. Afterwards cells were placed on ice. In a Stratalinker (Stratagene) they irradiated with 100 mJ/cm^2 of 254 nm UV light. UV negative controls were placed on ice as well, but not exposed to UV light. Subsequently, nuclear extracts were prepared as described in section 8.5.2. In the following 500 µg of nuclear protein extract were used per IP. All other steps were carried out strictly following Wang et al. (Wang, Tollervey et al. 2009). The adapters and primers used for the cDNA library preparation were taken from an Illumina TruSeq kit: RNA 5’ Adapter (RA5), part # 150132205, 5’ GUUCAGAGUUCUACAGUCCGACGAUC , RNA 3’ Adapter (RA3), part # 150132207, 5’ TGGAATTCTCGGGTGCCAAGG, RNA RT Primer (RTP), part # 15013981, 5’ AATGATACGGCGACCACCGAGATCTACACGTTCAGAGTTCTACAGTCCGA.

8.5.8 CLIP data analysis

Raw RNA-Seq reads were aligned to the mouse genome and visualized in the Integrative Genomics Viewer (IGV, Broad Institute). Data quality was assessed using the FastQC program (Babraham Bioinformatics).

8.5.9 Mass spectrometry

Essentially, SAF-A/hnRNP U IPs were conducted as described under the subsection "Immunoprecipitation". After the IP samples were run on a 4-12 % NuPAGE Bis-Tris acrylamide gel (Invitrogen). The gel was stained with SYPRO Ruby protein gel stain (Bio-Rad) following the supplier's instructions. Protein staining was visualized using a Gel Doc imaging system (Bio-Rad) and protein bands and IgG control bands were cut around SAF-A/hnRNP U's molecular weight (MW). The gel bands were cut into 1 mm cubes and submitted to the Dunn School of Pathology proteomics core facility (University of Oxford) where the samples were digested with trypsin and run on an electrospray ionization quadrupole time-of-flight (ESI-QUAD-TOF) mass spectrometer. The peptides were assigned to the *mus musculus* proteome using the Mascot server. For the detection of potential post-translationally modified sites an error tolerant Mascot search on was conducted on SAF-A/hnRNP U peptides (Hirosawa, Hoshida et al. 1993).

Alternatively, the samples were run only 5 mm into the gel and all protein was cut off the gel. Then samples were diced into 1 mm cubes and submitted to the Dunn School of Pathology proteomics core facility (University of Oxford). The raw data were submitted to the Central Proteomics Facility Pipeline (CPFP) (Trudgian, Thomas et al. 2010). A spectral index quantification (SINQ) meta-search for both IgG control and SAF-A/hnRNP U was used to retrieve the relative ratios of spectral counts [SC(SAF-A/hnRNP U)/ SC(IgG) and/or SC(IgG)/SC(SAF-A/hnRNP U)] with a false discovery rate (FDR) of 1 % and a maximum p-value of 0.01 % (Trudgian, Ridlova et al. 2011). Subsequently, the $\log_2$ of this ratio was taken to assess the enrichment or underrepresentation of a protein in the sample pool

relative to the IgG control. Proteins identified in the SAF-A/hnRNP U sample only were not included in this analysis and are listed separately.

Furthermore, three technical and biological replicates were analyzed using a power law global error model (plgem) R script written by Dr. David Trudgian. MIC normalized spectral indices (SIn) were used as input values. The quality of the underlying data was assessed using the power law global error model, a histogram plot of the residuals, a plot of residuals vs. rank and a normal Q-Q plot. The volcano plot was generated plotting the $\log_2$ mean fold change of the MIC SIn values versus the $-\log_{10}$ plgem p-value. The plot was used to identify proteins that were significantly enriched in the SAF-A/hnRNP U sample over the IgG control.

Later the fold change data were used to perform gross gene ontology (GO) analysis with MetaCore (Thomson Reuters), an integrated software suite for functional analysis of proteomic data.

8.5.10 Phos-tag™ sample and gel preparation

Nuclear extracts were prepared following the protocol described in Chapter 8.5.2. Buffers were supplemented with the phosphatase inhibitor PhosSTOP (Roche), or 10 mM *N*-Ethylmaleimide (NEM) and 10 µM of the cell-permeable proteasome inhibitor MG132 (CAS 133407-82-6, Calbiochem), respectively. Selected samples were incubated with alkaline phosphatase (Thermo Scientific FastAP EF0652) for 10 min at 37ºC that was heat inactivated through incubation at 75ºC for 5 min and addition of 50 mM EDTA.

PhosTag™ (Wako) gels were prepared following the basic protocol for SDS-PAGE gels (Chapter 8.5.4). The gel solution was supplemented with 10 – 100 µM Phos-tag™AAL-107 reagent, and 10 mM $MnCl_2$ were added to the separating gel. Gels were run at 5 – 15 mA. Subsequently, proteins were transferred onto a PVDF membrane following the manufacturer's instructions (12V, 20-25 min, semi-dry blotting). Afterwards, membranes

were incubated with anti-SAF-A/hnRNP U antibody and a corresponding secondary anti-rabbit secondary antibody.

8.6 General molecular biology techniques

8.6.1 RNA extraction

In most cases RNA was extracted using TRIzol Reagent (Invitrogen) following the manufacturer's instructions. Briefly, RNA was extracted from cellular pellets or directly from cell layers grown in the tissue culture dish. Before lysing the cells using the TRIzol Reagent (Invitrogen) the cells had been washed once with PBS. After 10 min of lysis chloroform was added to the extract, and the mixture was shaken vigorously for 15 s before leaving it to settle for 2 to 3 min at room temperature. Following this incubation step the different phases were separated by centrifugation for 15 min, at 12,000 g, at room temperature. The upper aqueous phase was transferred into a fresh tube and an equal volume of isopropanol was added. Subsequently, the RNA was left for precipitation on ice for 10 to 15 min. Then, the precipitate was spun down (12,000 g, 30 min, and 4 °C). The nucleic acid pellet was washed with 75 % (v/v) ethanol. Eventually, the RNA pellet was air-dried and re-suspended in RNase-free water.

Contaminating DNA was removed from the sample using the TURBO DNA-free (Invitrogen) reagent following the supplier's instructions. The obtained RNA samples were stored at -80 °C.

8.6.2 Complementary DNA synthesis

Complementary DNA (cDNA) was prepared from 2 µg isolated RNA. The reaction was primed with random pd$(N)_6$ hexamers (GE Healthcare) using SuperScript III Reverse Transcriptase (Invitrogen) according to the supplier's guidelines. The obtained cDNA was diluted with molecular biology grade ddH_2O and stored at -20 °C.

8.6.3 Quantitative real-time PCR

Quantitative real-time PCR (qPCR) was performed on cDNA templates using iQ SYBR Green Supermix (Bio-Rad) and 0.3 µM qPCR compatible forward and reverse primers (c.f.

Table 8.2). The reaction was run on a Chromo4 Real-Time Detector (Bio-Rad) and runs were analyzed with instrument's Opticon Monitor 3 (MJ Research Inc. Waltham, MA) software and Excel (Microsoft).

Table 8.2 – qPCR primes sets.

QPCR primer names and corresponding sequences are displayed in 5' to 3' orientation.

primer names	primer sequences (5' to 3')
Xist forward	GCAAGGAAGACAAAGGCTCAAAGAAT
Xist reverse	GGAGAGAGAACCAAATAGAGCAGAAT
β-actin forward	TGTTACCAACTGGGACGACA
β-actin reverse	CCATCACAATGCCTGTGGTA

8.7 Tissue culture

8.7.1 Cell culture media

The following media recipes used in this study were: EC10, prepared from Gibco DMEM (Invitrogen) supplemented with 10 % (v/v) fetal calf serum (FCS [biosera]), 100 µg/mL streptomycin-penicillin (Invitrogen), 2 mM L-glutamine (Invitrogen), 0.1 mM non-essential amino acids (Invitrogen) and 50 µM β-mercaptoethanol (Invitrogen), ESC medium from Gibco DMEM (Invitrogen) supplemented with 20 % (v/v) fetal calf serum (FCS) (biosera), 100 µg/mL streptomycin-penicillin (Invitrogen), 2 mM L-glutamine (Invitrogen), 0.1 mM non-essential amino acids (Invitrogen), 60 µM β-mercaptoethanol (Invitrogen) and 1 mL of leukemia inhibitory factor (LIF [prepared in house]), DMEM: F12 made up with 10 % (v/v) FCS (biosera), 100 µg/mL streptomycin-penicillin (Invitrogen), 2 mM L-glutamine (Invitrogen), 0.1 mM non-essential amino acids (Invitrogen) and 50 µM β-mercaptoethanol (Invitrogen), trophoblast stem cell (TSC) medium prepared with RPMI 1640 (Invitrogen)

supplemented with 20 % (v/v) FCS (biosera), 50 µg/mL streptomycin-penicillin (Invitrogen), 1 mM sodium pyruvate (Invitrogen), 2 mM L-glutamine, 100 µM β-mercaptoethanol (Invitrogen), 25 µg/mL fibroblast growth factor 4 (FGF4) and 1 µg/mL heparin, EC10 freezing medium made up of 50 % (v/v) FCS (biosera), 40 % (v/v) DMEM (Invitrogen) and 10 % (v/v) dimethylsulfoxide (DMSO [Sigma]) and ESC freezing medium contained 90 % (v/v) FCS (biosera) and 10 % (v/v) DMSO.

8.7.2 Freezing and thawing cells

Firstly, cells were washed with an excess of PBS and subsequently they were trypsinized with 0.05 % (w/v) Trypsin-EDTA (1x), Phenol Red (Invitrogen) supplemented with 2 % (v/v) chicken serum. Once the cells had detached form the culture dish surface, they were re-suspended with fresh medium and spun down in a MultiFuge 3L (Heraeus) for 3 min at 1,000 rpm. After aspirating the supernatant cells were re-suspended in the appropriate freezing medium and transferred into cryotubes (Nunc). The tubes were placed in Mr. Frosty Freezing Containers (Thermo Scientific) that were placed at -80 °C. Once frozen cells were stored in liquid nitrogen vapor phase.

To thaw up frozen cells, the cryovials were floated in a 33 or 37 °C water bath. After cells were thawed the DMSO (Sigma) content of the freezing medium was diluted down by adding normal growth medium to the cells. Once again the cells were spun down for 3 min at 1,000 rpm.

8.7.3 Feeder cell preparation

SNL feeder cells were expanded and inactivated by mitomycin C treatment. Thereby, SNLs were incubated for 2 h, at 37 °C and a 5% CO_2 atmosphere in EC10 supplemented with 10 µg/mL mitomycin C. Post incubation the cells were washed three times with an excess of PBS. Then the SNLs were trypsinized. Upon trypsinisation cells were counted and spun down. Eventually, cells were frozen down in suitable aliquots.

8.7.4 Lipofection of plasmid DNA

All lipofections were done using Lipofectamine 2000 Transfection Reagent (Invitrogen), mostly following the supplier's protocol. For this purpose cells were grown to 80 to 90 % confluency. Both plasmid DNA and the lipofection reagent were diluted in Opti-MEM I (Invitrogen) without serum and left incubating for 5 min at room temperature. Then the diluted DNA and Lipofectamine were combined, gently mixed and left to incubate another 20 min at room temperature. Post incubation the complexes were added to the medium in the tissue culture dish with the cells and gently mixed. After 16 to 20 h the medium was changed.

8.7.5 Kill curve preparation

To determine a cell line's sensitivity to a particular antibiotics (puromycin [Sigma] or Geneticin G418 [Invitrogen]) a kill curve was generated for the line in question. This was done by seeding the cells at equal density over a 6-well dish (Fisher). Once, the cells had grown to 50 % confluency the antibiotic of interest was added at various concentrations. Fresh medium with antibiotics was added on a daily basis and the cell death was monitored over several days. Eventually, the medium was aspirated from the plates and the wells were washed with PBS. Afterwards, the remaining cells were fixed and stained using a 6 % (v/v) glutaraldehyde, 0.5 % (w/v) crystalviolet solution. Upon completion of the fixation and staining procedure the wells were washed with PBS and an image was taken. An exemplary image for a G418 kill curve is shown in Figure 8.1.

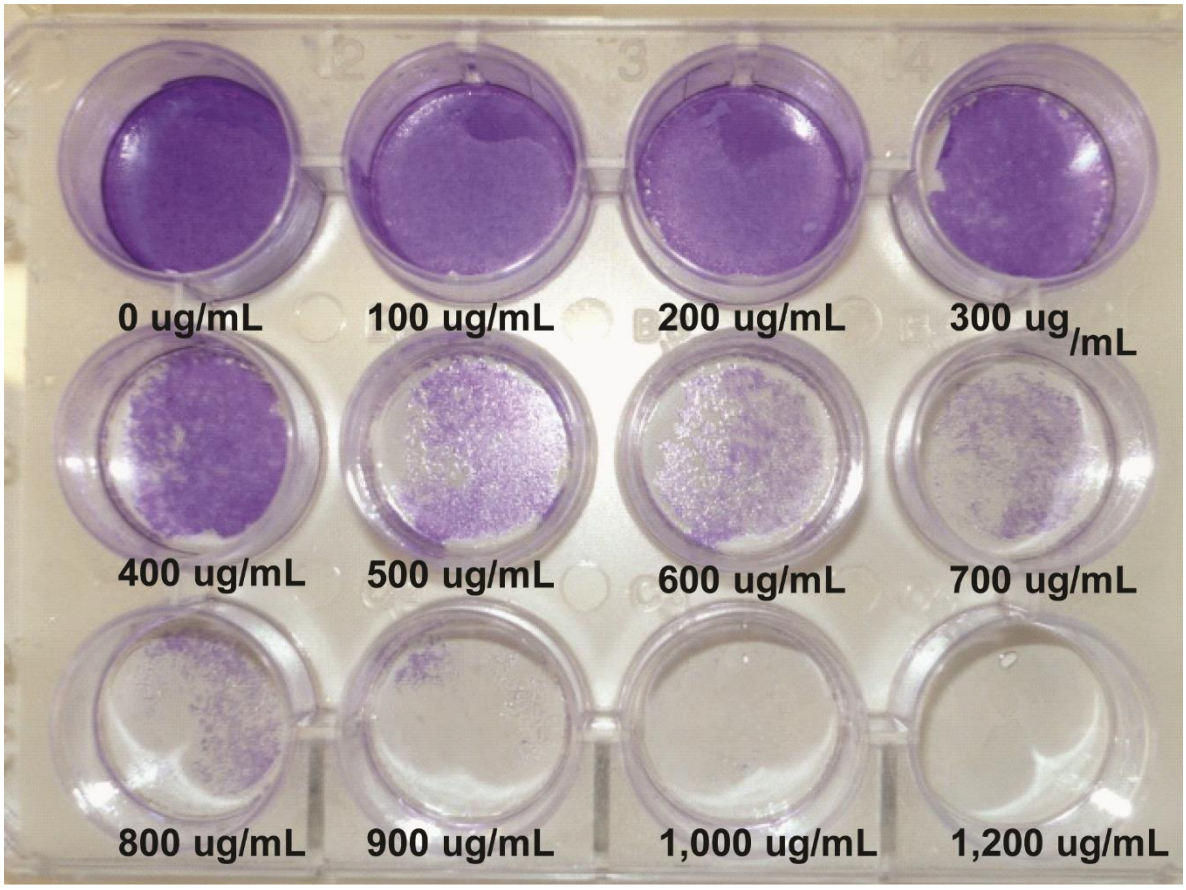


Figure 8.1 – Effect of Geneticin G418 on FSPE cell viability.

8.7.6 Cell lines

The below Table 8.3 displays the cell lines used, their sex and specific characteristics and special growth requirements.

Table 8.3 – Cell lines.

Table providing an overview on all cell lines used, their medium and special requirements, their sex, other characteristics and references to the published literature (Mak *et al.*2002 [Mak, Baxter et al. 2002], Alder *et al.* 2010 [Alder, Lavial et al. 2010], Yaffe *et al.* 1977 [Yaffe and Saxel 1977], Mermoud *et al.* 1999 [Mermoud, Costanzi et al. 1999], Olmsted *et al.* 1970 [Olmsted, Carlson et al. 1970], Tang *et al.* 2010 [Tang, Huntley et al. 2010], Penny *et al.* 1996 [Penny, Kay et al. 1996], Bodnar *et al.*1998 [Bodnar, Ouellette et al. 1998], McMahon and Bradley 1990 [McMahon and Bradley 1990], Pear *et al.* 1993 [Pear, Nolan et al. 1993]).

Cell line name	Medium and special requirements	XX/XY	Characteristics	Reference(s)
B1-TSC	TSC medium, PEFs, gelatinised plates (optional)	XX	TSC line derived from blastocysts from the cross (C57Bl6xCBA)F1 (XX) x PGK (XY). The use of PGK strain allows predefined polymorphisms on the X chromosome to be used to distinguish maternal (Xm) and paternal (Xp) alleles	Mak *et al.* 2002, Alder *et al.* 2010
C2C12	EC10	XX	*mus musculus* myoblast line derived from C3H (ATCC CRL-1772)	Yaffe *et al.* 1977
FSPE	EC10 medium	XX	fibroblast cell line derived from the ear tissue of *mus spretus* mouse	Mermoud *et al.* 1999
Neuro2a	EC10 medium	XX	neuronal amoeboid stem cell from *mus musculus* neuroblastoma (ATCC CLC-131)	Olmsted *et al.* 1970
P6D7	ESC medium, SNL-P feeders, gelatinised plates	XY	ESC line derived from 129 mice carrying a doxycycline-inducible Xist transgene on chromosome 17	Tang *et al.* 2010 (characterisation of parental cell line)
PGK12.1	ESC medium, geltinised plates	XX	ESC line derived from (PGKx129)F1(XY) x 129 (XX) cross; carries 1X chromosome from 129 strain and another from PGK strain of mice.	Penny *et al.* 1996
RPE-1	DMEM:F12	XX	Near-diploid *Homo sapiens* epithelial cell line derived from RPE-340 (ATCC CRL-4000)	Bodnar *et al.* 1998
SNLP 76/7-4	EC10 or ESC medium	XY	Fibroblast line derived from Sandos inbred mice, neomycin and puromycin resistant (ATCC SCRC-1050).	McMahon and Bradley 1990
HEK293T	EC10 medium	XXX	*Homo sapiens* embryonic kindney cell line constitutuively expressing SV40 large T-antigen (ATCC CRL-11268)	Pear *et al.* 1993
3E ESCs	ESC medium, SNL-P feeders, gelatinised plates	XY	Parental line of P6D7s	Tang *et al.* 2010

8.7.7 Stable cell line preparation and colony picking

Stable cell lines were made by either transfecting the cell line of interest with a plasmid encoding the protein of interest and a selectable resistance marker or by co-transfecting two plasmids, one coding for the protein of interest, the other one carrying the selectable marker. The selective marker whose working concentration had been determined in the kill curve experiment was applied about 24 h after transfection. After 7 to 12 days of selection single

clonal colonies were picked under the Leica MZ7.5 microscope and grown up in 96-well plates (Fisher). Once, individual clones had been expanded they were subjected to a screening for the expression of the protein of interest.

More specifically, the stable SAF-A/hnRNP U:hrGFP FSPE line was made by co-transfecting pCAGGS-mSAF-A/hnRNP U:hrGFP and pBS-loxP-puro in female *mus spretus* ear (FSPE) fibroblasts. After 11 days of selection with 4 µg/mL puromycin, individual clones were picked. Initially, selected clones were screened by conventional fluorescence microscopy. This was followed by Western blotting which was used to semi-quantitatively assess the protein's relative expression levels within the individual lines.

Subsequently, one of these clones was used to create a stable H2B:mCherry integration on top of the SAF-A/hnRNP U:hrGFP. This was done by co-transfecting a H2B:mCherry vector together with the G418 resistance carrying pGPT8. Afterwards cells were selected with 1.2 mg/mL G418 for up to 10 days. Conventional fluorescence microscopy was used to select for H2B:mCherry positive clones.

8.8 Microscopy techniques

8.8.1 Indirect immunofluorescence

Cells were grown on coverslips or slides and fixed either with 2 % (v/v) formaldehyde (Sigma) or 4 % (v/v) paraformaldehyde (Sigma) for 15 min. Alternatively, a mix of 50 % (v/v) methanol and 50 % (v/v) acetone or methanol at -20 °C was used to fix the cells for 5 min at room temperature. Fixation was followed by three washes with 1x PBS and 5 minutes permeabilization with 0.4 % (v/v) Triton-X 100. The coverslips were blocked with 0.2 % (w/v) PBS-fish gelatin (Sigma) solution and incubated with primary antibody at the appropriate dilution (Supplementary Table 1) in PBS-fish gelatin solution supplemented with 5 % (v/v) normal goat serum for 1 h. After three more washes with PBS-fish gelatin solution fluorophore-coupled secondary antibody was diluted in 5 % (v/v) normal goat serum

containing PBS-fish gelatin solution (Supplementary Table 1). The specimen and the secondary antibodies, which had been previously tested in IF experiments without primary antibodies, were incubated for another hour. Unbound secondary antibody was removed by washing the coverslips 3 times with PBS-fish gelatin. Subsequently, the specimens were mounted onto microscopy slides or covered up with coverslips with 4',6-diamidino-2-phenylindole (DAPI) VECTASHIELD Mounting Medium (vector laboratories) and sealed with nail varnish. Images were taken with an Axio Observer.Z1 microscope and processed with the AxioVision Rel. 4.8 software (Zeiss) and Adobe Photoshop.

8.8.2 Immuno-FISH for 3D-SIM and conventional fluorescence microscopy

C2C12 and RPE-1 cells transiently expressing SAF-A/hnRNP U:hrGFP were seeded on 18 × 18 mm borosilicate glass coverslips, number 1.5H (170 µm ± 5 µm thickness; Marienfeld Superior, Lauda-Königshofen, Germany). Then slides were washed twice with PBS and fixed in 2% PFA/PBS (v/v) for 10 min at room temperature. This was followed by a wash step with 0.05% PBS/Tween (v/v). Subsequently, cells were permeabilised in 0.5% Triton X-100/PBS (v/v) for 10 min at room temperature. The permeabilizing solution was washed off with 0.05% PBS/Tween (v/v). For hrGFP-Xist-FISH, cells were equilibrated in 2x saline-sodium citrate (SSC) buffer and incubated in 50% formamide/2x SSC at 4°C for 2 to 4 h. Labeled and denatured Xist-FISH probe was added, cells were mounted on slides, sealed with glue and samples were allowed to hybridize at 37°C overnight. Unbound probe was removed by washing slides three times with 2x SCC, three times with 0.05% Tween/2x SSC (SSCT) at 37°C. Probe detection was carried out in 2% BSA/0.5% fish skin gelatin (FSG)/4 x SSCT for 1 h at room temperature in a moist chamber. For subsequent immuno-staining, cells were equilibrated in 1 × PBST and blocked with 2% BSA/0.5% FSG/PBST for 1 h. Antibodies were diluted in blocking solution and incubated for 1 h followed by thorough washing with PBST. Afterward cells were post-fixed for 10 min in 4% FA/PBS. DNA was counterstained with 1 µg/ml DAPI or 0.25 µM SYTOX Green (Molecular Probes, Eugene,

OR, USA) in PBS for 10 min. Samples were mounted in Vectashield anti-fade mounting medium (Vector Laboratories, Burlingame, CA, USA) and sealed with nail varnish.

Immuno-RNA-FISH experiments had been carried out with and without RNase treatment. RNAse treatment eliminated Xist RNA foci.

8.8.3 3D-SIM microscopy

3D-SIM imaging was performed using a DeltaVision OMX 3V system (Applied Precision, GE Health Care) with a 63 ×/1.514 NA Plan Apo oil immersion objective (Olympus, Tokyo, Japan) and Cascade II:512 electron multiplying coupled charge device (EMCCD) cameras (Photometrics, Tucson, AZ, USA) and 405, 488 and 593 nm diode lasers. 256x256 images were taken in sequential mode and the light path was set to structured illumination (SI). Subsequently laser parameters were adjusted according to sample requirements. On every day of acquisition images of a control sample with single or double layers of triple-colored, 200 nm TetraSpeck (Invitrogen) beads and/or another control slide with triple-colored C127 cells (10 µM Click-IT® 5-ethynyl-2'-deoxyuridine [EdU] with CF™405M Azide [70%]; Alexa Fluor® 488 Azide [10%] and Alexa Fluor® 594 Azide [20%] [ThermoScientific]) were taken. Individual sample cells were selected based on tiling scans conducted with the microscope. Top and bottom of the specimen were set manually and images were acquired throughout the samples. Post image acquisition, the raw data were reconstructed in a workflow similar to the one described by Smeets *et al.* (Smeets, Markaki et al. 2014). Briefly, the data was computationally reconstructed using Wiener filter settings 0.002 and channel-specifically measured optical transfer functions (OTFs) using the softWoRX® 6.0 package (Applied Pecision). The OMX Editor software was used to obtain initial alignment parameters from control samples (i.e. TetraSpeck beads or C127 EdU control slide). Further image analysis was performed with the help of Dr. Lothar Schermelleh and Dr. David Pinto using the Fiji Is Just Image J software (http://fiji.sc/Fiji) and the corresponding plugins: Centroids were determined on 16-bit data, which were generated from 8-bit .log images using the SIMcheck

“Threshold and 16-bit conversion” utility (https://github.com/MicronOxford/SIMcheck). The analysis comprised the measurement of the distances between Xist foci (points of reference) to the nearest SAF-A foci (target points). Nearest neighbor analyses were performed using the TANGO plugin (Ollion, Cochennec et al. 2013).

8.8.4 FRAP and iFRAP

Live-cell imaging was performed on an Ultraview Spinning Disk (Perkin Elmer and Zeiss) microscope. Prior to imaging cells were plated on microscope-compatible 35 mm petri-dishes (µ-Dish ^{35mm,high}, ibidi). Whilst imaging cell culture dishes were kept in a ZLCS1-D35 microscope chamber (Tokai Hit) at 37 °C in a 5 % CO_2 atmosphere. FRAP and iFRAP data were acquired using the Volocity 3D Image Analysis Software (Perkin Elmer), and further analyzed with Fiji. Data were acquired and normalized as described by Dundr and Misteli (Dundr and Misteli 2003).

8.8.5 Long-term time-lapse imaging

For the long-term time lapse imaging cells were grown on borosilicate-coated and gelatinized chambered cover glass (Sigma). Right before imaging EC10 was exchanged against Opti-MEM I (Invitrogen) without phenol red supplemented with 2 % (v/v) FCS (biosera), 100 µg/mL streptomycin-penicillin (Invitrogen), 2 mM L-glutamine (Invitrogen), 0.1 mM non-essential amino acids (Invitrogen) and 50 µM β-mercaptoethanol (Invitrogen) and overlayed with mineral oil (Sigma). The cells were imaged using a Zeiss LSM510 with a heated stage and microscope chamber at 37 °C in a moisturized, 5 % CO_2 atmosphere. Up to 4 individual cells were followed over a period of up to 21 h by using the autofocus and a tracking macro developed by EMBL (Rabut and Ellenberg 2004). Z-stack images of each cell were taken at intervals of 5 to 6 min using less than 2 % laser power. The resulting image series were processed and analyzed with Fiji.

8.8.6 Fiji image analysis

The Fiji analysis software was used to adjust contrast and brightness, perform bleach correction via the “Exponential Fit” method, place scale bars and annotate time series. In addition it was used to determine fluorescence intensities in the regions of interest (ROIs) during FRAP analysis.

8.9 Flow cytometry

8.9.1 Propidium iodide staining for FACS

10^6 cells (FSPE) were harvested by trypsinization and spun down. After removal of the supernatant cells were washed with 10 mL PBS. Eventually the pellet was re-suspended in 3 mL PBS. Whilst vortexing 7 mL of ethanol were added and cells were left to fix on ice for 1 h. Then cells were spun down again and washed with PBS. The pellet was res-suspended in 1 mL PBS and cell clumps were separated by pipetting. Prior to FACS 10µL of 10mg/mL RNase A and 2µL of 2mg/mL propidium iodide were added to the cell suspension that was incubated in this solution for 30 min at 37°C in the dark.

8.9.2 DAPI staining for ImageStream IS100

Per sample $4x10^6$ FSPE cells were harvested by trypsinization. Samples were briefly equilibrated with PBS. Then cells were re-suspended in 1 mL PBS at room temperature. Subsequently, the full volume of cell suspension was transferred into ethanol whilst vortexing. Thereafter, cells were left in ethanol at -20 ºC for 5-15 min. Afterwards cells were pelleted by centrifugation and the ethanol supernatant was discarded. The cell pellet was loosened by tapping the tube and cells were rehydrated in 5 mL PBS at room temperature for 15 min. DAPI stock solution was diluted to 3 µM in staining buffer (100 mM Tris pH 7.4, 150 mM NaCl, 1 mM $CaCl_2$, 0.5 mM $MgCl_2$, 0.1 % NP-40). Cells were pelleted and re-suspended in 1 mL DAPI-staining solution per sample. The samples were stained 15 min at room temperature. Upon completion of the staining procedure cells were stored on ice and imaged using the ImageStream IS100 flow cytometer (Human Immunology Unit, John

Radcliffe Hospital, Oxford). ImageStream IS100 data were analyzed in the IDEAS software (Amnis).

8.10 Figure preparation

Figures were prepared using Adobe Illustrator CS3. Images were processed using Adobe Photoshop CS5, Fiji and Microsoft PowerPoint. To scale Venn diagrams were prepared using the online Venn Diagram Generator tool (http://jura.wi.mit.edu/bioc/tools/venn.php).

9. References

Abdelhaleem, M., L. Maltais and H. Wain (2003). "The human DDX and DHX gene families of putative RNA helicases." Genomics **81**(6): 618-622.

Abramoff, M. D., Magalhaes, P.J., Ram, S.J. (2004). "Image Processing with ImageJ." Biophotonics International **11**(7): 36-42.

Adam, S. A., T. Nakagawa, M. S. Swanson, T. K. Woodruff and G. Dreyfuss (1986). "mRNA polyadenylate-binding protein: gene isolation and sequencing and identification of a ribonucleoprotein consensus sequence." Mol Cell Biol **6**(8): 2932-2943.

Agnese, V., V. Bazan, F. P. Fiorentino, D. Fanale, G. Badalamenti, G. Colucci, V. Adamo, D. Santini and A. Russo (2007). "The role of Aurora-A inhibitors in cancer therapy." Ann Oncol **18 Suppl 6**: vi47-52.

Agostini, F., D. Cirillo, B. Bolognesi and G. G. Tartaglia (2013). "X-inactivation: quantitative predictions of protein interactions in the Xist network." Nucleic Acids Res **41**(1): e31.

Agrelo, R., A. Souabni, M. Novatchkova, C. Haslinger, M. Leeb, V. Komnenovic, H. Kishimoto, L. Gresh, T. Kohwi-Shigematsu, L. Kenner and A. Wutz (2009). "SATB1 defines the developmental context for gene silencing by Xist in lymphoma and embryonic cells." Dev Cell **16**(4): 507-516.

Alder, O., F. Lavial, A. Helness, E. Brookes, S. Pinho, A. Chandrashekran, P. Arnaud, A. Pombo, L. O'Neill and V. Azuara (2010). "Ring1B and Suv39h1 delineate distinct chromatin states at bivalent genes during early mouse lineage commitment." Development **137**(15): 2483-2492.

Altinok, A., D. Gonze, F. Levi and A. Goldbeter (2011). "An automaton model for the cell cycle." Interface Focus **1**(1): 36-47.

Andrews, S. "Babraham Bioinformatics." 2014, from http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

Arthold, S., A. Kurowski and A. Wutz (2011). "Mechanistic insights into chromosome-wide silencing in X inactivation." Hum Genet **130**(2): 295-305.

Ashworth, A., S. Rastan, R. Lovell-Badge and G. Kay (1991). "X-chromosome inactivation may explain the difference in viability of XO humans and mice." Nature **351**(6325): 406-408.

Aslanidis, C. and P. J. de Jong (1990). "Ligation-independent cloning of PCR products (LIC-PCR)." Nucleic Acids Research **18**(20): 6069-6074.

Balamotis, M. A., N. Tamberg, Y. J. Woo, J. Li, B. Davy, T. Kohwi-Shigematsu and Y. Kohwi (2012). "Satb1 ablation alters temporal expression of immediate early genes and reduces dendritic spine density during postnatal brain development." Mol Cell Biol **32**(2): 333-347.

Bailey, J. A., L. Carrel, A. Chakravarti and E. E. Eichler (2000). "Molecular evidence for a relationship between LINE-1 elements and X chromosome inactivation: the Lyon repeat hypothesis." Proc Natl Acad Sci U S A **97**(12): 6634-6639.

Barr, M. L. and E. G. Bertram (1949). "A morphological distinction between neurones of the male and female, and the behaviour of the nucleolar satellite during accelerated nucleoprotein synthesis." Nature **163**(4148): 676.

Bartke, T., M. Vermeulen, B. Xhemalce, S. C. Robson, M. Mann and T. Kouzarides (2010). "Nucleosome-interacting proteins regulated by DNA and histone methylation." Cell **143**(3): 470-484.

Baumann, C. and R. De La Fuente (2009). "ATRX marks the inactive X chromosome (Xi) in somatic cells and during imprinted X chromosome inactivation in trophoblast stem cells." Chromosoma **118**(2): 209-222.

Bennett, M., S. Pinol-Roma, D. Staknis, G. Dreyfuss and R. Reed (1992). "Differential binding of heterogeneous nuclear ribonucleoproteins to mRNA precursors prior to spliceosome assembly in vitro." Mol Cell Biol **12**(7): 3165-3175.

Berezney, R. (1991). "The nuclear matrix: a heuristic model for investigating genomic organization and function in the cell nucleus." J Cell Biochem **47**(2): 109-123.

Berezney, R. and D. S. Coffey (1974). "Identification of a nuclear protein matrix." Biochem Biophys Res Commun **60**(4): 1410-1417.

Berezney, R. and D. S. Coffey (1977). "Nuclear matrix. Isolation and characterization of a framework structure from rat liver nuclei." J Cell Biol **73**(3): 616-637.

Berglund, F. M. and P. R. Clarke (2009). "hnRNP-U is a specific DNA-dependent protein kinase substrate phosphorylated in response to DNA double-strand breaks." Biochem Biophys Res Commun **381**(1): 59-64.

Berletch, J. B., F. Yang, J. Xu, L. Carrel and C. M. Disteche (2011). "Genes that escape from X inactivation." Hum Genet **130**(2): 237-245.

Betzig, E., G. H. Patterson, R. Sougrat, O. W. Lindwasser, S. Olenych, J. S. Bonifacino, M. W. Davidson, J. Lippincott-Schwartz and H. F. Hess (2006). "Imaging intracellular fluorescent proteins at nanometer resolution." Science **313**(5793): 1642-1645.

Beyer, A. L., O. L. Miller, Jr. and S. L. McKnight (1980). "Ribonucleoprotein structure in nascent hnRNA is nonrandom and sequence-dependent." Cell **20**(1): 75-84.

Blewitt, M. E., A. V. Gendrel, Z. Pang, D. B. Sparrow, N. Whitelaw, J. M. Craig, A. Apedaile, D. J. Hilton, S. L. Dunwoodie, N. Brockdorff, G. F. Kay and E. Whitelaw (2008). "SmcHD1, containing a structural-maintenance-of-chromosomes hinge domain, has a critical role in X inactivation." Nat Genet **40**(5): 663-669.

Bode, J., T. Schlake, M. Rios-Ramirez, C. Mielke, M. Stengert, V. Kay and D. Klehr-Wirth (1995). "Scaffold/matrix-attached regions: structural properties creating transcriptionally active loci." Int Rev Cytol **162A**: 389-454.

Bodnar, A. G., M. Ouellette, M. Frolkis, S. E. Holt, C. P. Chiu, G. B. Morin, C. B. Harley, J. W. Shay, S. Lichtsteiner and W. E. Wright (1998). "Extension of life-span by introduction of telomerase into normal human cells." Science **279**(5349): 349-352.

Boyle, A. L., S. G. Ballard and D. C. Ward (1990). "Differential distribution of long and short interspersed element sequences in the mouse genome: chromosome karyotyping by fluorescence in situ hybridization." Proc Natl Acad Sci U S A **87**(19): 7757-7761.

Britton, S., C. Froment, P. Frit, B. Monsarrat, B. Salles and P. Calsou (2009). "Cell nonhomologous end joining capacity controls SAF-A phosphorylation by DNA-PK in response to DNA double-strand breaks inducers." Cell Cycle **8**(22): 3717-3722.

Brockdorff, N. (2002). "X-chromosome inactivation: closing in on proteins that bind Xist RNA." Trends Genet **18**(7): 352-358.

Brockdorff, N., A. Ashworth, G. F. Kay, P. Cooper, S. Smith, V. M. McCabe, D. P. Norris, G. D. Penny, D. Patel and S. Rastan (1991). "Conservation of position and exclusive expression of mouse Xist from the inactive X chromosome." Nature **351**(6324): 329-331.

Brockdorff, N., A. Ashworth, G. F. Kay, V. M. McCabe, D. P. Norris, P. J. Cooper, S. Swift and S. Rastan (1992). "The product of the mouse Xist gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus." Cell **71**(3): 515-526.

Brown, C. J. and S. E. Baldry (1996). "Evidence that heteronuclear proteins interact with XIST RNA in vitro." Somat Cell Mol Genet **22**(5): 403-417.

Brown, C. J., A. Ballabio, J. L. Rupert, R. G. Lafreniere, M. Grompe, R. Tonlorenzi and H. F. Willard (1991). "A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome." Nature **349**(6304): 38-44.

Brown, C. J., B. D. Hendrich, J. L. Rupert, R. G. Lafreniere, Y. Xing, J. Lawrence and H. F. Willard (1992). "The human XIST gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus." Cell **71**(3): 527-542.

Brown, C. J., R. G. Lafreniere, V. E. Powers, G. Sebastio, A. Ballabio, A. L. Pettigrew, D. H. Ledbetter, E. Levy, I. W. Craig and H. F. Willard (1991). "Localization of the X inactivation centre on the human X chromosome in Xq13." Nature **349**(6304): 82-84.

Burd, C. G., M. S. Swanson, M. Gorlach and G. Dreyfuss (1989). "Primary structures of the heterogeneous nuclear ribonucleoprotein A2, B1, and C2 proteins: a diversity of RNA

binding proteins is generated by small peptide inserts." Proc Natl Acad Sci U S A **86**(24): 9788-9792.

Burgoyne, P. S. and T. G. Baker (1981). "Oocyte depletion in XO mice and their XX sibs from 12 to 200 days post partum." J Reprod Fertil **61**(1): 207-212.

Burgoyne, P. S., E. P. Evans and K. Holland (1983). "XO monosomy is associated with reduced birthweight and lowered weight gain in the mouse." J Reprod Fertil **68**(2): 381-385.

Burgoyne, P. S. (1989). "Mammalian sex determination: thumbs down for zinc finger?" Nature **342**(6252): 860-862.

Burton, F. H., D. D. Loeb, C. F. Voliva, S. L. Martin, M. H. Edgell and C. A. Hutchison, 3rd (1986). "Conservation throughout mammalia and extensive protein-encoding capacity of the highly repeated DNA long interspersed sequence one." J Mol Biol **187**(2): 291-304.

Cadwell, K. and L. Coscoy (2005). "Ubiquitination on nonlysine residues by a viral E3 ubiquitin ligase." Science **309**(5731): 127-130.

Calabrese, J. M., W. Sun, L. Song, J. W. Mugford, L. Williams, D. Yee, J. Starmer, P. Mieczkowski, G. E. Crawford and T. Magnuson (2012). "Site-specific silencing of regulatory elements as a mechanism of X inactivation." Cell **151**(5): 951-963.

Caliebe, A., H. Y. Kroes, J. J. van der Smagt, J. I. Martin-Subero, H. Tonnies, R. van 't Slot, R. A. Nievelstein, H. Muhle, U. Stephani, K. Alfke, I. Stefanova, Y. Hellenbroich, G. Gillessen-Kaesbach, R. Hochstenbach, R. Siebert and M. Poot (2010). "Four patients with speech delay, seizures and variable corpus callosum thickness sharing a 0.440 Mb deletion in region 1q44 containing the HNRPU gene." Eur J Med Genet **53**(4): 179-185.

Cao, R., Y. Tsukada and Y. Zhang (2005). "Role of Bmi-1 and Ring1A in H2A ubiquitylation and Hox gene silencing." Mol Cell **20**(6): 845-854.

Cao, R., L. Wang, H. Wang, L. Xia, H. Erdjument-Bromage, P. Tempst, R. S. Jones and Y. Zhang (2002). "Role of histone H3 lysine 27 methylation in Polycomb-group silencing." Science **298**(5595): 1039-1043.

Carmena, M. and W. C. Earnshaw (2003). "The cellular geography of aurora kinases." Nat Rev Mol Cell Biol **4**(11): 842-854.

Carpenter, B., C. MacKay, A. Alnabulsi, M. MacKay, C. Telfer, W. T. Melvin and G. I. Murray (2006). "The roles of heterogeneous nuclear ribonucleoproteins in tumour development and progression." Biochim Biophys Acta **1765**(2): 85-100.

Carrel, L. and H. F. Willard (2005). "X-inactivation profile reveals extensive variability in X-linked gene expression in females." Nature **434**(7031): 400-404.

Cartegni, L., M. Maconi, E. Morandi, F. Cobianchi, S. Riva and G. Biamonti (1996). "hnRNP A1 selectively interacts through its Gly-rich domain with different RNA-binding proteins." J Mol Biol **259**(3): 337-348.

Cattanach, B. M. (1961). "A chemically-induced variegated-type position effect in the mouse." Z Vererbungsl **92**: 165-182.

Cattanach, B. M. and J. H. Isaacson (1967). "Controlling elements in the mouse X chromosome." Genetics **57**(2): 331-346.

Cattanach, B. M., J. N. Perez and C. E. Pollard (1970). "Controlling elements in the mouse X-chromosome. II. Location in the linkage map." Genet Res **15**(2): 183-195.

Cattanach, B. M., C. E. Pollard and J. N. Perez (1969). "Controlling elements in the mouse X-chromosome. I. Interaction with the X-linked genes." Genet Res **14**(3): 223-235.

Cerase, A., D. Smeets, Y. A. Tang, M. Gdula, F. Kraus, M. Spivakov, B. Moindrot, M. Leleu, A. Tattermusch, J. Demmerle, T. B. Nesterova, C. Green, A. P. Otte, L. Schermelleh and N. Brockdorff (2014). "Spatial separation of Xist RNA and polycomb proteins revealed by superresolution microscopy." Proc Natl Acad Sci U S A **111**(6): 2235-2240.

Chan, K. M., H. Zhang, L. Malureanu, J. van Deursen and Z. Zhang (2011). "Diverse factors are involved in maintaining X chromosome inactivation." Proc Natl Acad Sci U S A **108**(40): 16699-16704.

Charlesworth, B. (1978). "Model for evolution of Y chromosomes and dosage compensation." Proc Natl Acad Sci U S A **75**(11): 5618-5622.

Charlesworth, B. (1991). "The evolution of sex chromosomes." Science **251**(4997): 1030-1033.

Chaumeil, J., P. Le Baccon, A. Wutz and E. Heard (2006). "A novel role for Xist RNA in the formation of a repressive nuclear compartment into which genes are recruited when silenced." Genes & Development **20**(16): 2223-2237.

Chaumeil, J., P. Le Baccon, A. Wutz and E. Heard (2006). "A novel role for Xist RNA in the formation of a repressive nuclear compartment into which genes are recruited when silenced." Genes Dev **20**(16): 2223-2237.

Chow, J. C., C. Ciaudo, M. J. Fazzari, N. Mise, N. Servant, J. L. Glass, M. Attreed, P. Avner, A. Wutz, E. Barillot, J. M. Greally, O. Voinnet and E. Heard (2010). "LINE-1 activity in

facultative heterochromatin formation during X chromosome inactivation." Cell **141**(6): 956-969.

Chow, J. C., Z. Yen, S. M. Ziesche and C. J. Brown (2005). "Silencing of the mammalian X chromosome." Annu Rev Genomics Hum Genet **6**: 69-92.

Chureau, C., M. Prissette, A. Bourdet, V. Barbe, L. Cattolico, L. Jones, A. Eggen, P. Avner and L. Duret (2002). "Comparative sequence analysis of the X-inactivation center region in mouse, human, and bovine." Genome Res **12**(6): 894-908.

Civelekoglu-Scholey, G. and D. Cimini (2014). "Modelling chromosome dynamics in mitosis: a historical perspective on models of metaphase and anaphase in eukaryotic cells." Interface Focus **4**(3): 20130073.

Clemson, C. M., L. L. Hall, M. Byron, J. McNeil and J. B. Lawrence (2006). "The X chromosome is organized into a gene-rich outer rim and an internal core containing silenced nongenic sequences." Proc Natl Acad Sci U S A **103**(20): 7688-7693.

Clemson, C. M., J. A. McNeil, H. F. Willard and J. B. Lawrence (1996). "XIST RNA paints the inactive X chromosome at interphase: evidence for a novel RNA involved in nuclear/chromosome structure." J Cell Biol **132**(3): 259-275.

Cockerill, P. N. and W. T. Garrard (1986). "Chromosomal loop anchorage of the kappa immunoglobulin gene occurs next to the enhancer in a region containing topoisomerase II sites." Cell **44**(2): 273-282.

Costanzi, C. and J. R. Pehrson (1998). "Histone macroH2A1 is concentrated in the inactive X chromosome of female mammals." Nature **393**(6685): 599-601.

Craig, J. M. (2005). "Heterochromatin--many flavours, common themes." Bioessays **27**(1): 17-28.

Csankovszki, G., B. Panning, B. Bates, J. R. Pehrson and R. Jaenisch (1999). "Conditional deletion of Xist disrupts histone macroH2A localization but not maintenance of X inactivation." Nat Genet **22**(4): 323-324.

Davey, C. A., D. F. Sargent, K. Luger, A. W. Maeder and T. J. Richmond (2002). "Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 a resolution." J Mol Biol **319**(5): 1097-1113.

Davis, M., A. Hatzubai, J. S. Andersen, E. Ben-Shushan, G. Z. Fisher, A. Yaron, A. Bauskin, F. Mercurio, M. Mann and Y. Ben-Neriah (2002). "Pseudosubstrate regulation of the SCF(beta-TrCP) ubiquitin ligase by hnRNP-U." Genes Dev **16**(4): 439-451.

de Napoles, M., J. E. Mermoud, R. Wakao, Y. A. Tang, M. Endoh, R. Appanah, T. B. Nesterova, J. Silva, A. P. Otte, M. Vidal, H. Koseki and N. Brockdorff (2004). "Polycomb group proteins Ring1A/B link ubiquitylation of histone H2A to heritable gene silencing and X inactivation." Dev Cell **7**(5): 663-676.

Disteche, C. M., E. M. Eicher and S. A. Latt (1979). "Late replication in an X-autosome translocation in the mouse: correlation with genetic inactivation and evidence for selective effects during embryogenesis." Proc Natl Acad Sci U S A **76**(10): 5234-5238.

Dreyfuss, G., Y. D. Choi and S. A. Adam (1984). "Characterization of heterogeneous nuclear RNA-protein complexes in vivo with monoclonal antibodies." Mol Cell Biol **4**(6): 1104-1114.

Dreyfuss, G., V. N. Kim and N. Kataoka (2002). "Messenger-RNA-binding proteins and the messages they carry." Nat Rev Mol Cell Biol **3**(3): 195-205.

Dreyfuss, G., M. J. Matunis, S. Pinol-Roma and C. G. Burd (1993). "hnRNP proteins and the biogenesis of mRNA." Annu Rev Biochem **62**: 289-321.

Dundr, M. and T. Misteli (2003). "Measuring dynamics of nuclear proteins by photobleaching." Curr Protoc Cell Biol **Chapter 13**: Unit 13.5.

Duthie, S. M., T. B. Nesterova, E. J. Formstone, A. M. Keohane, B. M. Turner, S. M. Zakian and N. Brockdorff (1999). "Xist RNA exhibits a banded localization on the inactive X chromosome and is excluded from autosomal material in cis." Hum Mol Genet **8**(2): 195-204.

Ellis, N. and P. N. Goodfellow (1989). "The mammalian pseudoautosomal region." Trends Genet **5**(12): 406-410.

Engreitz, J. M., A. Pandya-Jones, P. McDonel, A. Shishkin, K. Sirokman, C. Surka, S. Kadri, J. Xing, A. Goren, E. S. Lander, K. Plath and M. Guttman (2013). "The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome." Science **341**(6147): 1237973.

Escamilla-Del-Arenal, M., S. T. da Rocha, C. G. Spruijt, O. Masui, O. Renaud, A. H. Smits, R. Margueron, M. Vermeulen and E. Heard (2013). "Cdyl, a New Partner of the Inactive X Chromosome and Potential Reader of H3K27me3 and H3K9me2." Mol Cell Biol **33**(24): 5005-5020.

Evans, K., S. Ott, A. Hansen, G. Koentges and L. Wernisch (2007). "A comparative study of S/MAR prediction tools." BMC Bioinformatics **8**: 71.

Fackelmayer, F. O. (2005). "A stable proteinaceous structure in the territory of inactive X chromosomes." J Biol Chem **280**(3): 1720-1723.

Fackelmayer, F. O., K. Dahm, A. Renz, U. Ramsperger and A. Richter (1994). "Nucleic-acid-binding properties of hnRNP-U/SAF-A, a nuclear-matrix protein which binds DNA and RNA in vivo and in vitro." Eur J Biochem **221**(2): 749-757.

Fackelmayer, F. O. and A. Richter (1994). "hnRNP-U/SAF-A is encoded by two differentially polyadenylated mRNAs in human cells." Biochim Biophys Acta **1217**(2): 232-234.

Fackelmayer, F. O. and A. Richter (1994). "Purification of two isoforms of hnRNP-U and characterization of their nucleic acid binding activity." Biochemistry **33**(34): 10416-10422.

Farcas, A. M., N. P. Blackledge, I. Sudbery, H. K. Long, J. F. McGouran, N. R. Rose, S. Lee, D. Sims, A. Cerase, T. W. Sheahan, H. Koseki, N. Brockdorff, C. P. Ponting, B. M. Kessler and R. J. Klose (2012). "KDM2B links the Polycomb Repressive Complex 1 (PRC1) to recognition of CpG islands." Elife **1**: e00205.

Ferguson-Smith, M. A. (1965). "Karyotype-Phenotype Correlations in Gonadal Dysgenesis and Their Bearing on the Pathogenesis of Malformations." J Med Genet **2**(2): 142-155.

Fey, P., B. Pflug and K. Jost (1988). "Self-rating of the global state of mood of depressive patients by a bright-dark-scale." Pharmacopsychiatry **21**(6): 410-411.

Finch, J. T. and A. Klug (1976). "Solenoidal model for superstructure in chromatin." Proc Natl Acad Sci U S A **73**(6): 1897-1901.

Fisher, E. M., P. Beer-Romero, L. G. Brown, A. Ridley, J. A. McNeil, J. B. Lawrence, H. F. Willard, F. R. Bieber and D. C. Page (1990). "Homologous ribosomal protein genes on the human X and Y chromosomes: escape from X inactivation and possible implications for Turner syndrome." Cell **63**(6): 1205-1218.

Fischle, W., Y. Wang, S. A. Jacobs, Y. Kim, C. D. Allis and S. Khorasanizadeh (2003). "Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by Polycomb and HP1 chromodomains." Genes Dev **17**(15): 1870-1881.

Fontes, A. (2013). "Cloning technologies." Methods Mol Biol **997**: 253-261.

Fu, D. and K. Collins (2007). "Purification of human telomerase complexes identifies factors involved in telomerase biogenesis and telomere length regulation." Mol Cell **28**(5): 773-785.

Gartler, S. M. and A. D. Riggs (1983). "Mammalian X-chromosome inactivation." Annu Rev Genet **17**: 155-190.

Gasser, S. M. and U. K. Laemmli (1986). "Cohabitation of scaffold binding regions with upstream/enhancer elements of three developmentally regulated genes of D. melanogaster." Cell **46**(4): 521-530.

Gehring, W. J. (1992). "The homeobox in perspective." Trends Biochem Sci **17**(8): 277-280.

Gibson, T. J., M. Seiler and R. A. Veitia (2013). "The transience of transient overexpression." Nat Methods **10**(8): 715-721.

Girod, P. A., D. Q. Nguyen, D. Calabrese, S. Puttini, M. Grandjean, D. Martinet, A. Regamey, D. Saugy, J. S. Beckmann, P. Bucher and N. Mermod (2007). "Genome-wide

prediction of matrix attachment regions that increase gene expression in mammalian cells." Nat Methods **4**(9): 747-753.

Glisovic, T., J. L. Bachorik, J. Yong and G. Dreyfuss (2008). "RNA-binding proteins and post-transcriptional gene regulation." FEBS Lett **582**(14): 1977-1986.

Gohring, F. and F. O. Fackelmayer (1997). "The scaffold/matrix attachment region binding protein hnRNP-U (SAF-A) is directly bound to chromosomal DNA in vivo: a chemical cross-linking study." Biochemistry **36**(27): 8276-8283.

Gohring, F., B. L. Schwab, P. Nicotera, M. Leist and F. O. Fackelmayer (1997). "The novel SAR-binding domain of scaffold attachment factor A (SAF-A) is a target in apoptotic nuclear breakdown." EMBO J **16**(24): 7361-7371.

Goldman, S., Spector (2010). Live Cell Imaging: A Laboratory Manual, Cold Spring Harbor Laboratory Press.

Goodfellow, P., B. Pym, T. Mohandas and L. J. Shapiro (1984). "The cell surface antigen locus, MIC2X, escapes X-inactivation." Am J Hum Genet **36**(4): 777-782.

Gustafsson, M. G. (2000). "Surpassing the lateral resolution limit by a factor of two using structured illumination microscopy." J Microsc **198**(Pt 2): 82-87.

Gustafsson, M. G. (2005). "Nonlinear structured-illumination microscopy: wide-field fluorescence imaging with theoretically unlimited resolution." Proc Natl Acad Sci U S A **102**(37): 13081-13086.

Hancock, R. (2000). "A new look at the nuclear matrix." Chromosoma **109**(4): 219-225.

Hafner, M., M. Landthaler, L. Burger, M. Khorshid, J. Hausser, P. Berninger, A. Rothballer, M. Ascano, Jr., A. C. Jungkamp, M. Munschauer, A. Ulrich, G. S. Wardle, S. Dewell, M. Zavolan and T. Tuschl (2010). "Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP." Cell **141**(1): 129-141.

Hall, L. L., M. Byron, G. Pageau and J. B. Lawrence (2009). "AURKB-mediated effects on chromatin regulate binding versus release of XIST RNA to the inactive chromosome." J Cell Biol **186**(4): 491-507.

Hall, L. L., D. M. Carone, A. V. Gomez, H. J. Kolpa, M. Byron, N. Mehta, F. O. Fackelmayer and J. B. Lawrence (2014). "Stable C0T-1 repeat RNA is abundant and is associated with euchromatic interphase chromosomes." Cell **156**(5): 907-919.

Hall, L. L. and J. B. Lawrence (2003). "The cell biology of a novel chromosomal RNA: chromosome painting by XIST/Xist RNA initiates a remodeling cascade." Semin Cell Dev Biol **14**(6): 369-378.

Han, S. P., Y. H. Tang and R. Smith (2010). "Functional diversity of the hnRNPs: past, present and perspectives." Biochem J **430**(3): 379-392.

Handel, M. A. (2004). "The XY body: a specialized meiotic chromatin domain." Exp Cell Res **296**(1): 57-63.

Hasegawa, Y., N. Brockdorff, S. Kawano, K. Tsutui, K. Tsutui and S. Nakagawa (2010). "The matrix protein hnRNP U is required for chromosomal localization of Xist RNA." Dev Cell **19**(3): 469-476.

Hasegawa, Y. and S. Nakagawa (2011). "Revisiting the function of nuclear scaffold/matrix binding proteins in X chromosome inactivation." RNA Biol **8**(5): 735-739.

Hauf, S., R. W. Cole, S. LaTerra, C. Zimmer, G. Schnapp, R. Walter, A. Heckel, J. van Meel, C. L. Rieder and J. M. Peters (2003). "The small molecule Hesperadin reveals a role for Aurora B in correcting kinetochore-microtubule attachment and in maintaining the spindle assembly checkpoint." J Cell Biol **161**(2): 281-294.

Heard, E., C. Rougeulle, D. Arnaud, P. Avner, C. D. Allis and D. L. Spector (2001). "Methylation of histone H3 at Lys-9 is an early mark on the X chromosome during X inactivation." Cell **107**(6): 727-738.

Heath, H., C. Ribeiro de Almeida, F. Sleutels, G. Dingjan, S. van de Nobelen, I. Jonkers, K. W. Ling, J. Gribnau, R. Renkawitz, F. Grosveld, R. W. Hendriks and N. Galjart (2008). "CTCF regulates cell cycle progression of alphabeta T cells in the thymus." EMBO J **27**(21): 2839-2850.

Hegde, M. L., S. Banerjee, P. M. Hegde, L. J. Bellot, T. K. Hazra, I. Boldogh and S. Mitra (2012). "Enhancement of NEIL1 protein-initiated oxidized DNA base excision repair by heterogeneous nuclear ribonucleoprotein U (hnRNP-U) via direct interaction." J Biol Chem **287**(41): 34202-34211.

Helbig, R. and F. O. Fackelmayer (2003). "Scaffold attachment factor A (SAF-A) is concentrated in inactive X chromosome territories through its RGG domain." Chromosoma **112**(4): 173-182.

Herrmann, F., M. Bossert, A. Schwander, E. Akgun and F. O. Fackelmayer (2004). "Arginine methylation of scaffold attachment factor A by heterogeneous nuclear ribonucleoprotein particle-associated PRMT1." J Biol Chem **279**(47): 48774-48779.

Hess, S. T., T. P. Girirajan and M. D. Mason (2006). "Ultra-high resolution imaging by fluorescence photoactivation localization microscopy." Biophys J **91**(11): 4258-4272.

Hirosawa, M., M. Hoshida, M. Ishikawa and T. Toya (1993). "MASCOT: multiple alignment system for protein sequences based on three-way dynamic programming." Comput Appl Biosci **9**(2): 161-167.

Hunt, P. A. (1991). "Survival of XO mouse fetuses: effect of parental origin of the X chromosome or uterine environment?" Development **111**(4): 1137-1141.

Huynh, K. D. and J. T. Lee (2001). "Imprinted X inactivation in eutherians: a model of gametic execution and zygotic relaxation." Curr Opin Cell Biol **13**(6): 690-697.

Jablonka, E. and M. J. Lamb (1990). "The evolution of heteromorphic sex chromosomes." Biol Rev Camb Philos Soc **65**(3): 249-276.

Jameson, B. A. and H. Wolf (1988). "The antigenic index: a novel algorithm for predicting antigenic determinants." Comput Appl Biosci **4**(1): 181-186.

Jansen, A. and K. J. Verstrepen (2011). "Nucleosome positioning in Saccharomyces cerevisiae." Microbiol Mol Biol Rev **75**(2): 301-320.

Jeon, Y. and J. T. Lee (2011). "YY1 tethers Xist RNA to the inactive X nucleation center." Cell **146**(1): 119-133.

Jeppesen, P. and B. M. Turner (1993). "The inactive X chromosome in female mammals is distinguished by a lack of histone H4 acetylation, a cytogenetic marker for gene expression." Cell **74**(2): 281-289.

Jonkers, I., K. Monkhorst, E. Rentmeester, J. A. Grootegoed, F. Grosveld and J. Gribnau (2008). "Xist RNA is confined to the nuclear territory of the silenced X chromosome throughout the cell cycle." Mol Cell Biol **28**(18): 5583-5594.

Jordan, I. K., I. B. Rogozin, G. V. Glazko and E. V. Koonin (2003). "Origin of a substantial fraction of human regulatory sequences from transposable elements." Trends Genet **19**(2): 68-72.

Kaneko, S., G. Li, J. Son, C. F. Xu, R. Margueron, T. A. Neubert and D. Reinberg (2010). "Phosphorylation of the PRC2 component Ezh2 is cell cycle-regulated and up-regulates its binding to ncRNA." Genes Dev **24**(23): 2615-2620.

Kanhere, A., K. Viiri, C. C. Araujo, J. Rasaiyaah, R. D. Bouwman, W. A. Whyte, C. F. Pereira, E. Brookes, K. Walker, G. W. Bell, A. Pombo, A. G. Fisher, R. A. Young and R. G. Jenner (2010). "Short RNAs are transcribed from repressed polycomb target genes and interact with polycomb repressive complex-2." Mol Cell **38**(5): 675-688.

Kawano, S., M. Miyaji, S. Ichiyasu, K. M. Tsutsui and K. Tsutsui (2010). "Regulation of DNA Topoisomerase IIbeta through RNA-dependent association with heterogeneous nuclear ribonucleoprotein U (hnRNP U)." J Biol Chem **285**(34): 26451-26460.

Keer, J. T., R. M. Hamvas, N. Brockdorff, D. Page, S. Rastan and S. D. Brown (1990). "Genetic mapping in the region of the mouse X-inactivation center." Genomics **7**(4): 566-572.

Kent, W. J., C. W. Sugnet, T. S. Furey, K. M. Roskin, T. H. Pringle, A. M. Zahler and D. Haussler (2002). "The human genome browser at UCSC." Genome Res **12**(6): 996-1006.

Kiledjian, M. and G. Dreyfuss (1992). "Primary structure and binding activity of the hnRNP U protein: binding RNA through RGG box." EMBO J **11**(7): 2655-2664.

Kim, M. K. and V. M. Nikodem (1999). "hnRNP U inhibits carboxy-terminal domain phosphorylation by TFIIH and represses RNA polymerase II elongation." Mol Cell Biol **19**(10): 6833-6844.

Kinoshita, E. and E. Kinoshita-Kikuta (2011). "Improved Phos-tag SDS-PAGE under neutral pH conditions for advanced protein phosphorylation profiling." Proteomics **11**(2): 319-323.

Kinoshita, E., E. Kinoshita-Kikuta and T. Koike (2012). "Phos-tag SDS-PAGE systems for phosphorylation profiling of proteins with a wide range of molecular masses under neutral pH conditions." Proteomics **12**(2): 192-202.

Kipp, M., F. Gohring, T. Ostendorp, C. M. van Drunen, R. van Driel, M. Przybylski and F. O. Fackelmayer (2000). "SAF-Box, a conserved protein domain that specifically recognizes scaffold attachment region DNA." Mol Cell Biol **20**(20): 7480-7489.

Kner, P., B. B. Chhun, E. R. Griffis, L. Winoto and M. G. Gustafsson (2009). "Super-resolution video microscopy of live cells by structured illumination." Nat Methods **6**(5): 339-342.

Kohlmaier, A., F. Savarese, M. Lachner, J. Martens, T. Jenuwein and A. Wutz (2004). "A chromosomal memory triggered by Xist regulates histone methylation in X inactivation." PLoS Biol **2**(7): E171.

Konig, J., K. Zarnack, G. Rot, T. Curk, M. Kayikci, B. Zupan, D. J. Turner, N. M. Luscombe and J. Ule (2010). "iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution." Nat Struct Mol Biol **17**(7): 909-915.

Konig, J., K. Zarnack, G. Rot, T. Curk, M. Kayikci, B. Zupan, D. J. Turner, N. M. Luscombe and J. Ule (2011). "iCLIP--transcriptome-wide mapping of protein-RNA interactions with individual nucleotide resolution." J Vis Exp (50): pii:2638.

Korenberg, J. R. and M. C. Rykowski (1988). "Human genome organization: Alu, lines, and the molecular structure of metaphase chromosome bands." Cell **53**(3): 391-400.

Krecic, A. M. and M. S. Swanson (1999). "hnRNP complexes: composition, structure, and function." Curr Opin Cell Biol **11**(3): 363-371.

Kucera, K. S., T. E. Reddy, F. Pauli, J. Gertz, J. E. Logan, R. M. Myers and H. F. Willard (2011). "Allele-specific distribution of RNA polymerase II on female X chromosomes." Hum Mol Genet **20**(20): 3964-3973.

Kukalev, A. S., I. B. Lobov, P. Percipalle and O. I. Podgornaya (2009). "SAF-A/hnRNP-U localization in interphase and metaphase." Cytogenet Genome Res **124**(3-4): 288-297.

Laemmli, U. K. (1970). "Cleavage of structural proteins during the assembly of the head of bacteriophage T4." Nature **227**(5259): 680-685.

Lee, J. T. and R. Jaenisch (1997). "Long-range cis effects of ectopic X-inactivation centres on a mouse autosome." Nature **386**(6622): 275-279.

Leppig, K. A., C. J. Brown, S. L. Bressler, K. Gustashaw, R. A. Pagon, H. F. Willard and C. M. Disteche (1993). "Mapping of the distal boundary of the X-inactivation center in a rearranged X chromosome from a female expressing XIST." Hum Mol Genet **2**(7): 883-887.

Lesage, B., J. Qian and M. Bollen (2011). "Spindle checkpoint silencing: PP1 tips the balance." Curr Biol **21**(21): R898-903.

Lieberman-Aiden, E., N. L. van Berkum, L. Williams, M. Imakaev, T. Ragoczy, A. Telling, I. Amit, B. R. Lajoie, P. J. Sabo, M. O. Dorschner, R. Sandstrom, B. Bernstein, M. A. Bender, M. Groudine, A. Gnirke, J. Stamatoyannopoulos, L. A. Mirny, E. S. Lander and J. Dekker (2009). "Comprehensive mapping of long-range interactions reveals folding principles of the human genome." Science **326**(5950): 289-293.

Liebich, I., J. Bode, M. Frisch and E. Wingender (2002). "S/MARt DB: a database on scaffold/matrix attached regions." Nucleic Acids Res **30**(1): 372-374.

Liebich, I., J. Bode, I. Reuter and E. Wingender (2002). "Evaluation of sequence motifs found in scaffold/matrix-attached regions (S/MARs)." Nucleic Acids Res **30**(15): 3433-3442.

Lin, H., V. Gupta, M. D. Vermilyea, F. Falciani, J. T. Lee, L. P. O'Neill and B. M. Turner (2007). "Dosage compensation in the mouse balances up-regulation and silencing of X-linked genes." PLoS Biol **5**(12): e326.

Lill, J. R. and I. E. Wertz (2014). "Toward understanding ubiquitin-modifying enzymes: from pharmacological targeting to proteomics." Trends Pharmacol Sci **35**(4): 187-207.

Lou, G., Y. Li, B. Chen, M. Chen, J. Chen, R. Liao, Y. Zhang, Y. Wang and D. Zhou (2007). "Functional analysis on the 5'-flanking region of human FXR gene in HepG2 cells." Gene **396**(2): 358-368.

Luger, K., A. W. Mader, R. K. Richmond, D. F. Sargent and T. J. Richmond (1997). "Crystal structure of the nucleosome core particle at 2.8 A resolution." Nature **389**(6648): 251-260.

Lynch, M. D., A. J. Smith, M. De Gobbi, M. Flenley, J. R. Hughes, D. Vernimmen, H. Ayyub, J. A. Sharpe, J. A. Sloane-Stanley, L. Sutherland, S. Meek, T. Burdon, R. J. Gibbons, D. Garrick and D. R. Higgs (2012). "An interspecies analysis reveals a key role for unmethylated CpG dinucleotides in vertebrate Polycomb complex recruitment." EMBO J **31**(2): 317-329.

Lyon, M. F. (1961). "Gene action in the X-chromosome of the mouse (Mus musculus L.)." Nature **190**: 372-373.

Lyon, M. F. (1998). "X-chromosome inactivation: a repeat hypothesis." Cytogenet Cell Genet **80**(1-4): 133-137.

Lyon, M. F. and S. G. Hawker (1973). "Reproductive lifespan in irradiated and unirradiated chromosomally XO mice." Genet Res **21**(2): 185-194.

Lyon, M. F., J. Zenthon, E. P. Evans, M. D. Burtenshaw, K. A. Wareham and E. D. Williams (1986). "Lack of inactivation of a mouse X-linked gene physically separated from the inactivation centre." J Embryol Exp Morphol **97**: 75-85.

Ma, N., S. Matsunaga, A. Morimoto, G. Sakashita, T. Urano, S. Uchiyama and K. Fukui (2011). "The nuclear scaffold protein SAF-A is required for kinetochore-microtubule

attachment and contributes to the targeting of Aurora-A to mitotic spindles." J Cell Sci **124**(Pt 3): 394-404.

Maenner, S., M. Blaud, L. Fouillen, A. Savoye, V. Marchand, A. Dubois, S. Sanglier-Cianferani, A. Van Dorsselaer, P. Clerc, P. Avner, A. Visvikis and C. Branlant (2010). "2-D structure of the A region of Xist RNA and its implication for PRC2 association." PLoS Biol **8**(1): e1000276.

Mak, W., J. Baxter, J. Silva, A. E. Newall, A. P. Otte and N. Brockdorff (2002). "Mitotically stable association of polycomb group proteins eed and enx1 with the inactive x chromosome in trophoblast stem cells." Curr Biol **12**(12): 1016-1020.

Mao, S., E. Sendler, R. J. Goodrich, R. Hauser and S. A. Krawetz (2014). "A comparison of sperm RNA-seq methods." Syst Biol Reprod Med: 1-8.

Margueron, R., N. Justin, K. Ohno, M. L. Sharpe, J. Son, W. J. Drury, 3rd, P. Voigt, S. R. Martin, W. R. Taylor, V. De Marco, V. Pirrotta, D. Reinberg and S. J. Gamblin (2009). "Role of the polycomb protein EED in the propagation of repressive histone marks." Nature **461**(7265): 762-767.

Markaki, Y., D. Smeets, S. Fiedler, V. J. Schmid, L. Schermelleh, T. Cremer and M. Cremer (2012). "The potential of 3D-FISH and super-resolution structured illumination microscopy for studies of 3D nuclear architecture: 3D structured illumination microscopy of defined chromosomal structures visualized by 3D (immuno)-FISH opens new perspectives for studies of nuclear architecture." Bioessays **34**(5): 412-426.

Markham, K., Y. Bai and G. Schmitt-Ulms (2007). "Co-immunoprecipitations revisited: an update on experimental concepts and their implementation for sensitive interactome investigations of endogenous proteins." Anal Bioanal Chem **389**(2): 461-473.

Martens, J. H., R. J. O'Sullivan, U. Braunschweig, S. Opravil, M. Radolf, P. Steinlein and T. Jenuwein (2005). "The profile of repeat-associated histone lysine methylation states in the mouse epigenome." EMBO J **24**(4): 800-812.

Martens, J. H., M. Verlaan, E. Kalkhoven, J. C. Dorsman and A. Zantema (2002). "Scaffold/matrix attachment region elements interact with a p300-scaffold attachment factor A complex and are bound by acetylated nucleosomes." Mol Cell Biol **22**(8): 2598-2606.

Martin, S. L. and F. D. Bushman (2001). "Nucleic acid chaperone activity of the ORF1 protein from the mouse LINE-1 retrotransposon." Mol Cell Biol **21**(2): 467-475.

Mattern, K. A., B. M. Humbel, A. O. Muijsers, L. de Jong and R. van Driel (1996). "hnRNP proteins and B23 are the major proteins of the internal nuclear matrix of HeLa S3 cells." J Cell Biochem **62**(2): 275-289.

Mattern, K. A., I. van der Kraan, W. Schul, L. de Jong and R. van Driel (1999). "Spatial organization of four hnRNP proteins in relation to sites of transcription, to nuclear speckles, and to each other in interphase nuclei and nuclear matrices of HeLa cells." Exp Cell Res **246**(2): 461-470.

Matunis, E. L., M. J. Matunis and G. Dreyfuss (1993). "Association of individual hnRNP proteins and snRNPs with nascent transcripts." J Cell Biol **121**(2): 219-228.

McMahon, A. P. and A. Bradley (1990). "The Wnt-1 (int-1) proto-oncogene is required for development of a large region of the mouse brain." Cell **62**(6): 1073-1085.

Meisenheimer, K. M. and T. H. Koch (1997). "Photocross-linking of nucleic acids to associated proteins." Crit Rev Biochem Mol Biol **32**(2): 101-140.

Mermoud, J. E., C. Costanzi, J. R. Pehrson and N. Brockdorff (1999). "Histone macroH2A1.2 relocates to the inactive X chromosome after initiation and propagation of X-inactivation." J Cell Biol **147**(7): 1399-1408.

Min, J., Y. Zhang and R. M. Xu (2003). "Structural basis for specific binding of Polycomb chromodomain to histone H3 methylated at Lys 27." Genes Dev **17**(15): 1823-1828.

Mirkovitch, J., M. E. Mirault and U. K. Laemmli (1984). "Organization of the higher-order chromatin loop: specific DNA attachment sites on nuclear scaffold." Cell **39**(1): 223-232.

Monk, M. and M. Harper (1978). "X-chromosome activity in preimplantation mouse embryos from XX and XO mothers." J Embryol Exp Morphol **46**: 53-64.

Muller, H. J. (1964). "The Relation of Recombination to Mutational Advance." Mutat Res **106**: 2-9.

Munkley, J., N. A. Copeland, V. Moignard, J. R. Knight, E. Greaves, S. A. Ramsbottom, M. E. Pownall, J. Southgate, J. F. Ainscough and D. Coverley (2011). "Cyclin E is recruited to the nuclear matrix during differentiation, but is not recruited in cancer cells." Nucleic Acids Res **39**(7): 2671-2677.

Nakamura, K. C., H. Kameda, Y. Koshimizu, Y. Yanagawa and T. Kaneko (2008). "Production and histological application of affinity-purified antibodies to heat-denatured green fluorescent protein." J Histochem Cytochem **56**(7): 647-657.

Naughton, C., D. Sproul, C. Hamilton and N. Gilbert (2010). "Analysis of active and inactive X chromosome architecture reveals the independent organization of 30 nm and large-scale chromatin structures." Mol Cell **40**(3): 397-409.

Ng, K., N. Daigle, A. Bancaud, T. Ohhata, P. Humphreys, R. Walker, J. Ellenberg and A. Wutz (2011). "A system for imaging the regulatory noncoding Xist RNA in living mouse embryonic stem cells." Mol Biol Cell **22**(14): 2634-2645.

Norris, D. P., D. Patel, G. F. Kay, G. D. Penny, N. Brockdorff, S. A. Sheardown and S. Rastan (1994). "Evidence that random and imprinted Xist expression is controlled by preemptive methylation." Cell **77**(1): 41-51.

Nozawa, R. S., K. Nagao, K. T. Igami, S. Shibata, N. Shirai, N. Nozaki, T. Sado, H. Kimura and C. Obuse (2013). "Human inactive X chromosome is compacted through a PRC2-independent SMCHD1-HBiX1 pathway." Nat Struct Mol Biol **20**(5): 566-573.

Obrdlik, A., A. Kukalev, E. Louvet, A. K. Farrants, L. Caputo and P. Percipalle (2008). "The histone acetyltransferase PCAF associates with actin and hnRNP U for RNA polymerase II transcription." Mol Cell Biol **28**(20): 6342-6357.

Ohno, S., W. D. Kaplan and R. Kinosita (1959). "Formation of the sex chromatin by a single X-chromosome in liver cells of Rattus norvegicus." Exp Cell Res **18**: 415-418.

Okamoto, I., A. P. Otte, C. D. Allis, D. Reinberg and E. Heard (2004). "Epigenetic dynamics of imprinted X inactivation during early mouse development." Science **303**(5658): 644-649.

Olmsted, J. B., K. Carlson, R. Klebe, F. Ruddle and J. Rosenbaum (1970). "Isolation of microtubule protein from cultured mouse neuroblastoma cells." Proc Natl Acad Sci U S A **65**(1): 129-136.

Ollion, J., J. Cochennec, F. Loll, C. Escude and T. Boudier (2013). "TANGO: a generic tool for high-throughput 3D image analysis for studying nuclear organization." Bioinformatics **29**(14): 1840-1841.

Page, D. C., R. Mosher, E. M. Simpson, E. M. Fisher, G. Mardon, J. Pollack, B. McGillivray, A. de la Chapelle and L. G. Brown (1987). "The sex-determining region of the human Y chromosome encodes a finger protein." Cell **51**(6): 1091-1104.

Palmer, E. and T. Freeman (2004). "Investigation into the use of C- and N-terminal GFP fusion proteins for subcellular localization studies using reverse transfection microarrays." Comp Funct Genomics **5**(4): 342-353.

Palmer, M. S., A. H. Sinclair, P. Berta, N. A. Ellis, P. N. Goodfellow, N. E. Abbas and M. Fellous (1989). "Genetic evidence that ZFY is not the testis-determining factor." Nature **342**(6252): 937-939.

Pavelka, N., M. Pelizzola, C. Vizzardelli, M. Capozzoli, A. Splendiani, F. Granucci and P. Ricciardi-Castagnoli (2004). "A power law global error model for the identification of differentially expressed genes in microarray data." BMC Bioinformatics **5**: 203.

Payer, B. and J. T. Lee (2008). "X chromosome dosage compensation: how mammals keep the balance." Annu Rev Genet **42**: 733-772.

Pear, W. S., G. P. Nolan, M. L. Scott and D. Baltimore (1993). "Production of high-titer helper-free retroviruses by transient transfection." Proc Natl Acad Sci U S A **90**(18): 8392-8396.

Pederson, T. (1998). "Thinking about a nuclear matrix." J Mol Biol **277**(2): 147-159.

Pederson, T. (2000). "Half a century of "the nuclear matrix"." Mol Biol Cell **11**(3): 799-805.

Penny, G. D., G. F. Kay, S. A. Sheardown, S. Rastan and N. Brockdorff (1996). "Requirement for Xist in X chromosome inactivation." Nature **379**(6561): 131-137.

Pinol-Roma, S., Y. D. Choi, M. J. Matunis and G. Dreyfuss (1988). "Immunopurification of heterogeneous nuclear ribonucleoprotein particles reveals an assortment of RNA-binding proteins." Genes Dev **2**(2): 215-227.

Plath, K., J. Fang, S. K. Mlynarczyk-Evans, R. Cao, K. A. Worringer, H. Wang, C. C. de la Cruz, A. P. Otte, B. Panning and Y. Zhang (2003). "Role of histone H3 lysine 27 methylation in X inactivation." Science **300**(5616): 131-135.

Ponting, C., J. Schultz and P. Bork (1997). "SPRY domains in ryanodine receptors (Ca(2+)-release channels)." Trends Biochem Sci **22**(6): 193-194.

Posch, M., G. A. Khoudoli, S. Swift, E. M. King, J. G. Deluca and J. R. Swedlow (2010). "Sds22 regulates aurora B activity and microtubule-kinetochore interactions at mitosis." J Cell Biol **191**(1): 61-74.

Pullirsch, D., R. Hartel, H. Kishimoto, M. Leeb, G. Steiner and A. Wutz (2010). "The Trithorax group protein Ash2l and Saf-A are recruited to the inactive X chromosome at the onset of stable X inactivation." Development **137**(6): 935-943.

Pullirsch, D. and M. F. Jantsch (2010). "Proteome diversification by adenosine to inosine RNA editing." RNA Biol **7**(2): 205-212.

Purbowasito, W., C. Suda, T. Yokomine, M. Zubair, T. Sado, K. Tsutsui and H. Sasaki (2004). "Large-scale identification and mapping of nuclear matrix-attachment regions in the distal imprinted domain of mouse chromosome 7." DNA Res **11**(6): 391-407.

Rabinovitch, P. S., M. Kubbies, Y. C. Chen, D. Schindler and H. Hoehn (1988). "BrdU-Hoechst flow cytometry: a unique tool for quantitative cell cycle analysis." Experimental cell research **174**(2): 309-318.

Rabut, G. and J. Ellenberg (2004). "Automatic real-time three-dimensional cell tracking by fluorescence microscopy." J Microsc **216**(Pt 2): 131-137.

Rastan, S. (1982). "Timing of X-chromosome inactivation in postimplantation mouse embryos." J Embryol Exp Morphol **71**: 11-24.

Rastan, S. (1983). "Non-random X-chromosome inactivation in mouse X-autosome translocation embryos--location of the inactivation centre." J Embryol Exp Morphol **78**: 1-22.

Rastan, S. and S. D. Brown (1990). "The search for the mouse X-chromosome inactivation centre." Genet Res **56**(2-3): 99-106.

Rastan, S. and E. J. Robertson (1985). "X-chromosome deletions in embryo-derived (EK) cell lines associated with lack of X-chromosome inactivation." J Embryol Exp Morphol **90**: 379-388.

Rice, W. R. (1987). "Genetic hitchhiking and the evolution of reduced genetic activity of the Y sex chromosome." Genetics **116**(1): 161-167.

Riggs, A. D., J. Singer-Sam and D. H. Keith (1985). "Methylation of the PGK promoter region and an enhancer way-station model for X-chromosome inactivation." Prog Clin Biol Res **198**: 211-222.

Rinn, J. L. and J. Ule (2014). "'Oming in on RNA-protein interactions." Genome Biol **15**(1): 401.

Roshon, M. J. and H. E. Ruley (2005). "Hypomorphic mutation in hnRNP U results in post-implantation lethality." Transgenic Res **14**(2): 179-192.

Romig, H., F. O. Fackelmayer, A. Renz, U. Ramsperger and A. Richter (1992). "Characterization of SAF-A, a novel nuclear DNA binding protein from HeLa cells with high affinity for nuclear matrix/scaffold attachment DNA elements." EMBO J **11**(9): 3431-3440.

Roshon, M. J. and H. E. Ruley (2005). "Hypomorphic mutation in hnRNP U results in post-implantation lethality." Transgenic Research **14**(2): 179-192.

Ross, M. T., D. V. Grafham, A. J. Coffey, S. Scherer, K. McLay, D. Muzny, M. Platzer, G. R. Howell, C. Burrows, C. P. Bird, A. Frankish, F. L. Lovell, K. L. Howe, J. L. Ashurst, R. S. Fulton, R. Sudbrak, G. Wen, M. C. Jones, M. E. Hurles, T. D. Andrews, C. E. Scott, S. Searle, J. Ramser, A. Whittaker, R. Deadman, N. P. Carter, S. E. Hunt, R. Chen, A. Cree, P.

Gunaratne, P. Havlak, A. Hodgson, M. L. Metzker, S. Richards, G. Scott, D. Steffen, E. Sodergren, D. A. Wheeler, K. C. Worley, R. Ainscough, K. D. Ambrose, M. A. Ansari-Lari, S. Aradhya, R. I. Ashwell, A. K. Babbage, C. L. Bagguley, A. Ballabio, R. Banerjee, G. E. Barker, K. F. Barlow, I. P. Barrett, K. N. Bates, D. M. Beare, H. Beasley, O. Beasley, A. Beck, G. Bethel, K. Blechschmidt, N. Brady, S. Bray-Allen, A. M. Bridgeman, A. J. Brown, M. J. Brown, D. Bonnin, E. A. Bruford, C. Buhay, P. Burch, D. Burford, J. Burgess, W. Burrill, J. Burton, J. M. Bye, C. Carder, L. Carrel, J. Chako, J. C. Chapman, D. Chavez, E. Chen, G. Chen, Y. Chen, Z. Chen, C. Chinault, A. Ciccodicola, S. Y. Clark, G. Clarke, C. M. Clee, S. Clegg, K. Clerc-Blankenburg, K. Clifford, V. Cobley, C. G. Cole, J. S. Conquer, N. Corby, R. E. Connor, R. David, J. Davies, C. Davis, J. Davis, O. Delgado, D. Deshazo, P. Dhami, Y. Ding, H. Dinh, S. Dodsworth, H. Draper, S. Dugan-Rocha, A. Dunham, M. Dunn, K. J. Durbin, I. Dutta, T. Eades, M. Ellwood, A. Emery-Cohen, H. Errington, K. L. Evans, L. Faulkner, F. Francis, J. Frankland, A. E. Fraser, P. Galgoczy, J. Gilbert, R. Gill, G. Glockner, S. G. Gregory, S. Gribble, C. Griffiths, R. Grocock, Y. Gu, R. Gwilliam, C. Hamilton, E. A. Hart, A. Hawes, P. D. Heath, K. Heitmann, S. Hennig, J. Hernandez, B. Hinzmann, S. Ho, M. Hoffs, P. J. Howden, E. J. Huckle, J. Hume, P. J. Hunt, A. R. Hunt, J. Isherwood, L. Jacob, D. Johnson, S. Jones, P. J. de Jong, S. S. Joseph, S. Keenan, S. Kelly, J. K. Kershaw, Z. Khan, P. Kioschis, S. Klages, A. J. Knights, A. Kosiura, C. Kovar-Smith, G. K. Laird, C. Langford, S. Lawlor, M. Leversha, L. Lewis, W. Liu, C. Lloyd, D. M. Lloyd, H. Loulseged, J. E. Loveland, J. D. Lovell, R. Lozado, J. Lu, R. Lyne, J. Ma, M. Maheshwari, L. H. Matthews, J. McDowall, S. McLaren, A. McMurray, P. Meidl, T. Meitinger, S. Milne, G. Miner, S. L. Mistry, M. Morgan, S. Morris, I. Muller, J. C. Mullikin, N. Nguyen, G. Nordsiek, G. Nyakatura,

C. N. O'Dell, G. Okwuonu, S. Palmer, R. Pandian, D. Parker, J. Parrish, S. Pasternak, D. Patel, A. V. Pearce, D. M. Pearson, S. E. Pelan, L. Perez, K. M. Porter, Y. Ramsey, K. Reichwald, S. Rhodes, K. A. Ridler, D. Schlessinger, M. G. Schueler, H. K. Sehra, C. Shaw-Smith, H. Shen, E. M. Sheridan, R. Shownkeen, C. D. Skuce, M. L. Smith, E. C. Sotheran, H. E. Steingruber, C. A. Steward, R. Storey, R. M. Swann, D. Swarbreck, P. E. Tabor, S. Taudien, T. Taylor, B. Teague, K. Thomas, A. Thorpe, K. Timms, A. Tracey, S. Trevanion, A. C. Tromans, M. d'Urso, D. Verduzco, D. Villasana, L. Waldron, M. Wall, Q. Wang, J. Warren, G. L. Warry, X. Wei, A. West, S. L. Whitehead, M. N. Whiteley, J. E. Wilkinson, D. L. Willey, G. Williams, L. Williams, A. Williamson, H. Williamson, L. Wilming, R. L. Woodmansey, P. W. Wray, J. Yen, J. Zhang, J. Zhou, H. Zoghbi, S. Zorilla, D. Buck, R. Reinhardt, A. Poustka, A. Rosenthal, H. Lehrach, A. Meindl, P. J. Minx, L. W. Hillier, H. F. Willard, R. K. Wilson, R. H. Waterston, C. M. Rice, M. Vaudin, A. Coulson, D. L. Nelson, G. Weinstock, J. E. Sulston, R. Durbin, T. Hubbard, R. A. Gibbs, S. Beck, J. Rogers and D. R. Bentley (2005). "The DNA sequence of the human X chromosome." Nature **434**(7031): 325-337.

Rossbach, O., L. H. Hung, E. Khrameeva, S. Schreiner, J. Konig, T. Curk, B. Zupan, J. Ule, M. S. Gelfand and A. Bindereif (2014). "Crosslinking-immunoprecipitation (iCLIP) analysis reveals global regulatory roles of hnRNP L." RNA Biol **11**(2): 146-155.

Rothbart, S. B. and B. D. Strahl (2014). "Interpreting the language of histone and DNA modifications." Biochim Biophys Acta **1839**(8): 627-643.

Royce-Tolland, M. E., A. A. Andersen, H. R. Koyfman, D. J. Talbot, A. Wutz, I. D. Tonks, G. F. Kay and B. Panning (2010). "The A-repeat links ASF/SF2-dependent Xist RNA processing with random choice during X inactivation." Nat Struct Mol Biol **17**(8): 948-954.

Russell, L. B. (1963). "Mammalian X-chromosome action: inactivation limited in spread and region of origin." Science **140**(3570): 976-978.

Russell, L. B. and N. L. Cacheiro (1978). "The use of mouse X-autosome translocations in the study of X-inactivation pathways and nonrandomness." Basic Life Sci **12**: 393-416.

Schermelleh, L., R. Heintzmann and H. Leonhardt (2010). "A guide to super-resolution fluorescence microscopy." J Cell Biol **190**(2): 165-175.

Schoeftner, S., A. K. Sengupta, S. Kubicek, K. Mechtler, L. Spahn, H. Koseki, T. Jenuwein and A. Wutz (2006). "Recruitment of PRC1 function at the initiation of X inactivation independent of PRC2 and silencing." EMBO J **25**(13): 3110-3122.

Schuettengruber, B., D. Chourrout, M. Vervoort, B. Leblanc and G. Cavalli (2007). "Genome regulation by polycomb and trithorax proteins." Cell **128**(4): 735-745.

Sei, E. and N. K. Conrad (2014). "UV cross-linking of interacting RNA and protein in cultured cells." Methods Enzymol **539**: 53-66.

Shnyreva, M., D. S. Schullery, H. Suzuki, Y. Higaki and K. Bomsztyk (2000). "Interaction of two multifunctional proteins. Heterogeneous nuclear ribonucleoprotein K and Y-box-binding protein." J Biol Chem **275**(20): 15498-15503.

Silva, J., W. Mak, I. Zvetkova, R. Appanah, T. B. Nesterova, Z. Webster, A. H. Peters, T. Jenuwein, A. P. Otte and N. Brockdorff (2003). "Establishment of histone h3 methylation on the inactive X chromosome requires transient recruitment of Eed-Enx1 polycomb group complexes." Dev Cell **4**(4): 481-495.

Silva, J., W. Mak, I. Zvetkova, R. Appanah, T. B. Nesterova, Z. Webster, A. H. F. M. Peters, T. Jenuwein, A. P. Otte and N. Brockdorff (2003). "Establishment of histone h3 methylation on the inactive X chromosome requires transient recruitment of Eed-Enx1 polycomb group complexes." Developmental Cell **4**(4): 481-495.

Simon, M. D., S. F. Pinter, R. Fang, K. Sarma, M. Rutenberg-Schoenberg, S. K. Bowman, B. A. Kesner, V. K. Maier, R. E. Kingston and J. T. Lee (2013). "High-resolution Xist binding maps reveal two-step spreading during X-chromosome inactivation." Nature **504**(7480): 465-469.

Smeets, D., Y. Markaki, V. J. Schmid, F. Kraus, A. Tattermusch, A. Cerase, M. Sterr, S. Fiedler, J. Demmerle, J. Popken, H. Leonhardt, N. Brockdorff, T. Cremer, L. Schermelleh and M. Cremer (2014). "Three-dimensional super-resolution microscopy of the inactive X chromosome territory reveals a collapse of its active nuclear compartment harboring distinct Xist RNA foci." Epigenetics Chromatin **7**: 8.

Smith, K. P., M. Byron, C. M. Clemson and J. B. Lawrence (2004). "Ubiquitinated proteins including uH2A on the human and mouse inactive X chromosome: enrichment in gene rich bands." Chromosoma **113**(6): 324-335.

Spitzer, J., M. Hafner, M. Landthaler, M. Ascano, T. Farazi, G. Wardle, J. Nusbaum, M. Khorshid, L. Burger, M. Zavolan and T. Tuschl (2014). "PAR-CLIP (Photoactivatable Ribonucleoside-Enhanced Crosslinking and Immunoprecipitation): a step-by-step protocol to the transcriptome-wide identification of binding sites of RNA-binding proteins." Methods Enzymol **539**: 113-161.

Spraggon, L., T. Dudnakova, J. Slight, O. Lustig-Yariv, J. Cotterell, N. Hastie and C. Miles (2007). "hnRNP-U directly interacts with WT1 and modulates WT1 transcriptional activation." Oncogene **26**(10): 1484-1491.

Sprague, B. L. and J. G. McNally (2005). "FRAP analysis of binding: proper and fitting." Trends Cell Biol **15**(2): 84-91.

Steitz, J. A. and R. Kamen (1981). "Arrangement of 30S heterogeneous nuclear ribonucleoprotein on polyoma virus late nuclear transcripts." Mol Cell Biol **1**(1): 21-34.

Stuurman, N., A. M. Meijne, A. J. van der Pol, L. de Jong, R. van Driel and J. van Renswoude (1990). "The nuclear matrix from cells of different origin. Evidence for a common set of matrix proteins." J Biol Chem **265**(10): 5460-5465.

Swanson, M. S. and G. Dreyfuss (1988). "Classification and purification of proteins of heterogeneous nuclear ribonucleoprotein particles by RNA-binding specificities." Mol Cell Biol **8**(5): 2237-2241.

Sydnes, M. O. and M. Isobe (2010). "Tautomycin's interactions with protein phosphatase 1." Chem Asian J **5**(3): 410-420.

Takagi, N., O. Sugawara and M. Sasaki (1982). "Regional and temporal changes in the pattern of X-chromosome replication during the early post-implantation development of the female mouse." Chromosoma **85**(2): 275-286.

Takagi, N., M. Sugimoto, S. Yamaguchi, M. Ito, S. S. Tan and M. Okabe (2002). "Nonrandom X chromosome inactivation in mouse embryos carrying Searle's T(X;16)16H

translocation visualized using X-linked LACZ and GFP transgenes." Cytogenet Genome Res **99**(1-4): 52-58.

Talbert, P. B. and S. Henikoff (2006). "Spreading of silent chromatin: inaction at a distance." Nat Rev Genet **7**(10): 793-803.

Tang, Y. A., D. Huntley, G. Montana, A. Cerase, T. B. Nesterova and N. Brockdorff (2010). "Efficiency of Xist-mediated silencing on autosomes is linked to chromosomal domain organisation." Epigenetics Chromatin **3**(1): 10.

Taniura, H. and K. Yoshikawa (2002). "Necdin interacts with the ribonucleoprotein hnRNP U in the nuclear matrix." J Cell Biochem **84**(3): 545-555.

Tatchell, K., V. Makrantoni, M. J. Stark and L. C. Robinson (2011). "Temperature-sensitive ipl1-2/Aurora B mutation is suppressed by mutations in TOR complex 1 via the Glc7/PP1 phosphatase." Proc Natl Acad Sci U S A **108**(10): 3994-3999.

Taverna, S. D., H. Li, A. J. Ruthenburg, C. D. Allis and D. J. Patel (2007). "How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers." Nat Struct Mol Biol **14**(11): 1025-1040.

Taylor, J. and S. Woodcock (2015). "A Perspective on the Future of High-Throughput RNAi Screening: Will CRISPR Cut Out the Competition or Can RNAi Help Guide the Way?" J Biomol Screen.

Thoma, F., T. Koller and A. Klug (1979). "Involvement of histone H1 in the organization of the nucleosome and of the salt-dependent superstructures of chromatin." J Cell Biol **83**(2 Pt 1): 403-427.

Thorvaldsen, J. L., R. I. Verona and M. S. Bartolomei (2006). "X-tra! X-tra! News from the mouse X chromosome." Dev Biol **298**(2): 344-353.

Trojer, P. and D. Reinberg (2007). "Facultative heterochromatin: is there a distinctive molecular signature?" Molecular Cell **28**(1): 1-13.

Trudgian, D. C., G. Ridlova, R. Fischer, M. M. Mackeen, N. Ternette, O. Acuto, B. M. Kessler and B. Thomas (2011). "Comparative evaluation of label-free SINQ normalized spectral index quantitation in the central proteomics facilities pipeline." Proteomics **11**(14): 2790-2797.

Trudgian, D. C., B. Thomas, S. J. McGowan, B. M. Kessler, M. Salek and O. Acuto (2010). "CPFP: a central proteomics facilities pipeline." Bioinformatics **26**(8): 1131-1132.

Tuerk, C. and L. Gold (1990). "Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase." Science **249**(4968): 505-510.

Ule, J., K. Jensen, A. Mele and R. B. Darnell (2005). "CLIP: a method for identifying protein-RNA interaction sites in living cells." Methods **37**(4): 376-386.

Urlaub, H., K. Hartmuth and R. Luhrmann (2002). "A two-tracked approach to analyze RNA-protein crosslinking sites in native, nonlabeled small nuclear ribonucleoprotein particles." Methods **26**(2): 170-181.

Verschure, P. J., I. Van Der Kraan, J. M. Enserink, M. J. Mone, E. M. Manders and R. Van Driel (2002). "Large-scale chromatin organization and the localization of proteins involved in gene expression in human cells." J Histochem Cytochem **50**(10): 1303-1312.

Vizlin-Hodzic, D., H. Johansson, J. Ryme, T. Simonsson and S. Simonsson (2011). "SAF-A has a role in transcriptional regulation of Oct4 in ES cells through promoter binding." Cell Reprogram **13**(1): 13-27.

Vizlin-Hodzic, D., R. Runnberg, J. Ryme, S. Simonsson and T. Simonsson (2011). "SAF-A forms a complex with BRG1 and both components are required for RNA polymerase II mediated transcription." PLoS One **6**(12): e28049.

Wako. "Phos-tag PAGE guidebook." Retrieved 25.07.15, 2015, from http://www.wako-chem.co.jp/english/labchem/journals/phos-tag_GB2013/pdf/Phos-tag.pdf.

Wang, H., L. Wang, H. Erdjument-Bromage, M. Vidal, P. Tempst, R. S. Jones and Y. Zhang (2004). "Role of histone H2A ubiquitination in Polycomb silencing." Nature **431**(7010): 873-878.

Wang, X., R. A. Herr, W. J. Chua, L. Lybarger, E. J. Wiertz and T. H. Hansen (2007). "Ubiquitination of serine, threonine, or lysine residues on the cytoplasmic tail can induce ERAD of MHC-I by viral E3 ligase mK3." J Cell Biol **177**(4): 613-624.

Wang, Z., J. Tollervey, M. Briese, D. Turner and J. Ule (2009). "CLIP: construction of cDNA libraries for high-throughput sequencing from RNAs cross-linked to proteins in vivo." Methods **48**(3): 287-293.

Westphal, V., S. O. Rizzoli, M. A. Lauterbach, D. Kamin, R. Jahn and S. W. Hell (2008). "Video-rate far-field optical nanoscopy dissects synaptic vesicle movement." Science **320**(5873): 246-249.

Williams, K. R. and W. H. Konigsberg (1991). "Identification of amino acid residues at interface of protein-nucleic acid complexes by photochemical cross-linking." Methods Enzymol **208**: 516-539.

Willig, K. I., S. O. Rizzoli, V. Westphal, R. Jahn and S. W. Hell (2006). "STED microscopy reveals that synaptotagmin remains clustered after synaptic vesicle exocytosis." Nature **440**(7086): 935-939.

Wilson, R. H. and D. Coverley (2013). "Relationship between DNA replication and the nuclear matrix." Genes Cells **18**(1): 17-31.

Wutz, A. and R. Jaenisch (2000). "A shift from reversible to irreversible X inactivation is triggered during ES cell differentiation." Mol Cell **5**(4): 695-705.

Wutz, A., T. P. Rasmussen and R. Jaenisch (2002). "Chromosomal silencing and localization are mediated by different domains of Xist RNA." Nat Genet **30**(2): 167-174.

Xiao, R., P. Tang, B. Yang, J. Huang, Y. Zhou, C. Shao, H. Li, H. Sun, Y. Zhang and X. D. Fu (2012). "Nuclear matrix factor hnRNP U/SAF-A exerts a global control of alternative splicing by regulating U2 snRNP maturation." Mol Cell **45**(5): 656-668.

Xu, C., C. Bian, W. Yang, M. Galka, H. Ouyang, C. Chen, W. Qiu, H. Liu, A. E. Jones, F. MacKenzie, P. Pan, S. S. Li, H. Wang and J. Min (2010). "Binding of different histone marks

differentially regulates the activity and specificity of polycomb repressive complex 2 (PRC2)." Proc Natl Acad Sci U S A **107**(45): 19266-19271.

Yaffe, D. and O. Saxel (1977). "A myogenic cell line with altered serum requirements for differentiation." Differentiation **7**(3): 159-166.

Ye, J., N. Beetz, S. O'Keeffe, J. C. Tapia, L. Macpherson, W. V. Chen, R. Bassel-Duby, E. N. Olson and T. Maniatis (2015). "hnRNP U protein is required for normal pre-mRNA splicing and postnatal heart development and function." Proc Natl Acad Sci U S A **112**(23): E3020-3029.

Yugami, M., Y. Kabe, Y. Yamaguchi, T. Wada and H. Handa (2007). "hnRNP-U enhances the expression of specific genes by stabilizing mRNA." FEBS Lett **581**(1): 1-7.

Zacksenhaus, E. and R. Sheinin (1990). "Molecular cloning, primary structure and expression of the human X linked A1S9 gene cDNA which complements the ts A1S9 mouse L cell defect in DNA replication." EMBO J **9**(9): 2923-2929.

Zhang, L., K. Zhang, R. Prandl and F. Schoffl (2004). "Detecting DNA-binding of proteins in vivo by UV-crosslinking and immunoprecipitation." Biochem Biophys Res Commun **322**(3): 705-711.

Zhang, L. F., K. D. Huynh and J. T. Lee (2007). "Perinucleolar targeting of the inactive X during S phase: evidence for a role in the maintenance of silencing." Cell **129**(4): 693-706.

Zhang, S., B. Schlott, M. Gorlach and F. Grosse (2004). "DNA-dependent protein kinase (DNA-PK) phosphorylates nuclear DNA helicase II/RNA helicase A and hnRNP proteins in an RNA-dependent manner." Nucleic Acids Res **32**(1): 1-10.

Zhao, J., B. K. Sun, J. A. Erwin, J. J. Song and J. T. Lee (2008). "Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome." Science **322**(5902): 750-756.

Zinn, A. R., S. L. Bressler, P. Beer-Romero, D. A. Adler, V. M. Chapman, D. C. Page and C. M. Disteche (1991). "Inactivation of the Rps4 gene on the mouse X chromosome." Genomics **11**(4): 1097-1101.