Author:
Raele-Rolfe, Renata A

Title:
Investigating the functions of scaffold attachment factor B (SAFB) proteins following cellular stress

Investigating the functions of scaffold attachment factor B (SAFB) proteins following cellular stress

Renata Almeida Raele

A dissertation submitted to the University of Bristol in accordance with the requirements for the award of the degree of Doctor of Philosophy in the Faculty of Health Sciences, Bristol Medical School.

August 2020.

Abstract

Scaffold Attachment Factor B1 (SAFB1) and SAFB2 are RNA and DNA binding proteins with roles in gene expression, DNA repair and the stress response. SAFB1 was identified first and has been the most intensely studied and the high degree of similarity between the two proteins has led to difficulty in elucidating their individual roles. In this study I have used advanced genetic techniques and SAFB1 and SAFB2 specific antibodies to discriminate the functions of the individual family members. Specifically, experiments have been conducted to investigate: (i) the specific roles of SAFB1 and SAFB2 proteins in the stress response,(ii) how SAFB protein function is regulated by post-translational modification; (iii) what the SAFB protein binding partners. Immunocytochemistry using SAFB1 and SAFB2 specific antibodies revealed for the first time that SAFB2 is recruited to nuclear stress bodies (nSBs) following heat stress (HS), co-localising with HSF1 in 85% of cells displaying canonical HSF1+ nSBs, whereas SAFB1 co-localised with HSF1 in only 6.7% of these cells. SAFB2 also co-localised with the HS induced lncSATIII RNA. SAFB2 knockdown in SH-SY5Y cells led to an increase in the expression lncSATIII transcripts following heat stress, an effect that was not seen with either SAFB1 knockdown or double SAFB1/SAFB2 knockdown. These findings suggest that SAFB2 rather than SAFB1 as previously proposed plays an important in nSB formation and regulation of the stress response. Global methylation inhibition strongly decreased the formation of canonical and secondary HSF1 nSBs following HS, increased large SAFB1 aggregate formation, and had no effect on SAFB2 distribution. Meanwhile, a PRMT5 inhibitor (inhibition of symmetric di-methyl arginine) significantly increased the formation of canonical HSF1 nSBs but produced no effect on the distribution of any SAFB family member. Moreover, SAFB1 and SAFB2 mutants containing impaired sites of SUMOylation and arginine methylation showed a delay in their recruitment to HSF1 nSBs during HS. Finally, tandem mass tagging mass spectrometry proteomic analysis of proteins co-immunoprecipitated with SAFB1 and SAFB2 was used to identify novel interacting partners for SAFB1 and SAFB2 proteins. A number of novel interaction partners were identified and as our laboratory is interested in miRNA function particular interest was raised by the observation that both SAFB1 and SAFB2 both co-immunoprecipitated with Ago2. Ago2 is key regulator of miRNA mediated silencing and is a novel interacting partner of SAFB proteins not previously reported in the literature. Furthermore, Ingenuity Pathway Analysis (IPA) was performed to explore the cellular processes regulated by SAFB and its interaction partners. Analyses also showed that the interaction partners of both SAFB1 and SAFB2 proteins were involved in miRNA processing, eIF2/ISR signalling, the unfolded protein response and in Huntington's Disease signalling, and these findings are discussed more fully in the thesis. Together, these data reinforce a key role of SAFB1 and SAFB2 in regulating critical cellular processes such as miRNA processing, cellular adaptation to stress and neurological disease development.

Acknowledgements

This work was supported by the grant DPE-Doutorado Pleno no Exterior from the Coordination for the Improvement of Higher Education Personnel (CAPES), an educational division of the Brazilian Government.

I wish to thank the many people who helped me make this possible, people who gave me the gift of their knowledge, time, patience, and love.

First and foremost, I wish to express my sincere appreciation to my supervisor and co-supervisor, Professor James Uney and Professor Liang-Fong Wong, for giving me this opportunity to pursue a PhD under their supervision. Many thanks to Prf. James Uney for providing guidance and support throughout my journey. It has been a fantastic four years of research that I will never forget.

A special thanks go to Dr Helen Scott for her or her continuous and unwavering support, without it, I would not be here. Her kindness and patience when answering my questions meant a lot to me.

I am indebted to all my colleagues in the UWOC group: Dr Nicola Buckner, Dr Mamdouh Allahyani, Dr Jalilah Idris and Dr Stephanie Wallis. They have walked the same path as me and were with me during my time at UoB. I will never forget the laughs during endless coffee times. I wish to thank all the people whose assistance was valuable in completing this project: Dr. Oscar Cordero-Llana, Dr Mahmoud Khazim, Dr Hadil Al Ahdal, Dr Sian Baker, Miss Andriana Gialeli and Mr Ben Exley. I`d like to include in this Dr Miriam Durazo Barba, who`s immense contributions were crucial to the completion of this final document. Thank you also to the Biomedical Science Building B and C floor technical team, especially Mr Doug Owen, and to the professional team in the Dorothy Hodgkin Building, especially Mr Paul Newcomb, for providing valuable laboratory technical assistance.

I am thankful to Professor Mathias Altmeyer at the University of Zurich for kindly supplying the SAFB1 deletion constructs used in my research and Dr Lucy Crompton from the School of Biochemistry, University of Bristol, for the primary human fibroblasts. Moreover, I would like to pay my special regards to Dr Stephen Cross, from the Wolfson Bioimaging Facility of UoB, for all his time and patience spent with me at the Opera microscope/imaging analyses. Without his expertise, part of this would not be possible. Many thanks to Dr Phill Lewis and his proteomic analyses expertise. 4.

On a personal note, I want to express my deepest gratitude to my Mum, for her unconditional love and for all the concessions she made to let me go. Without her support, this work would not be possible. Mãe, você tem a minha admiração pela sua coragem em me deixar partir. Serei eternamente grata por ter você ao meu lado.

My heart will always be with Elisa Ribeiro, without whom I would not even have started.

Finally, my special regards go to my husband, Benjamin Rolfe, who has been the light of my days for the past four years. He kept me going on, and this work would not have been possible without his support and coffee making skills. I love you.

Author`s Declaration

I declare that the work in this dissertation was carried out in accordance with the requirements of the *University's Regulations and Code of Practice for Research Degree Programmes* and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others is indicated as such. Any views expressed in the dissertation are those of the author.

SIGNED: ..... DATE:.....

Table of Contents

List of Figures

List of Abbreviations

ATM	Ataxia-telangiectasia-mutated kinase
ATP	Adenosine triphosphate
ATR	ATM- and Rad3-Related Kinase
ATXN	Ataxin gene/ataxin RNA
BARD1	BRCA1 associated RING domain 1
BBC3	Bcl-2-binding component 3
BCA	Bicinchoninic acid assay
BIP	HSP70 family protein 5
BRCA1	Breast cancer type 1 susceptibility protein 1
BRG1	Brahma-related gene-1
CALR	Calreticulin
CAPES	Coordination for the Improvement of Higher Education Personnel
CBP	CREB-binding protein
CCT	T-complex protein 1 subunits
CDKN2A	Cyclin-dependent kinase inhibitor 2A
cDNA	Complementary DNA
ChIRP-MS	Comprehensive identification of RNA-binding proteins by mass spectrometry
CLBs	Cross-linked beads
CLSM	Confocal laser scanning microscopy
CLTC	Clathrin heavy chain
CNS	Central nervous system
Co-IP	Co-Immunoprecipitation
COS	**C**V-1 (simian) in **O**rigin, **S**V40 genetic material Cell Line (Monkey Kidney fibroblast-like cell line)
CREB	cAMP response element-binding protein
CRISPR	Clustered regularly interspaced short palindromic repeats
CT	Threshold cycle number
DAVID	Database for Annotation, Visualisation and Integrated Discovery
DDR	DNA damage response
DDX	DEAD-box helicase
DEAD	Asp-glu-ala-asp amino acid sequence containing proteins
DGCR8	Non-catalytic subunit of the Microprocessor Protein Complex
DMEM	Dulbecco's Modified Eagle Medium
DMP	Dimethyl pimelimidate dihydrochloride
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleoside triphosphate
DSB	Double Strand Break
DTT	Dithiothreitol
E.R.	Endoplasmic reticulum
EBI	European Bioinformatics Institute
ECL	Enhanced chemiluminescence
EDTA	Ethylenediaminetetraacetic acid
EGFP	Enhanced green fluorescent protein
EGFR	Epidermal growth factor receptor
EIF	Eukaryotic Initiation Factor

ELAVL	ELAV like RNA binding protein
ER	Oestrogen Receptor
ERE	Oestrogen response elements
ERH	Enhancer of Rudimentary Homologue
EZH2	Enhancer of zeste homolog 2
FAS	Fas cell surface death receptor
FBS	Foetal bovine serum
FDR	False discovery rate
FISH	Fluorescent *In Situ* Hybridisation
FMR1	FMRP translational regulator
FMRP	fragile X mental retardation protein
FUS	Fused in sarcoma
FXR	Farnesoid X Receptor
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase
GAR	Glycine-Arginine-Rich
GLI	Glycine
GO	Gene ontology
GSK3203591	PRMT5 Inhibitor (GSK591)
H2AX	Histone 2 AX
HAP	HnRNP A1 associated protein
HAT	Histone acetyltransferase
HB	Hydrogen-bonds
HDAC	histone deacetylase
HEK293T	Homo sapiens embryonic kidney cell line
HeLa	Henrietta Lacks` human cervical cancer cells
HET	Hsp27 ERE TATA-binding protein
hIPCs	Human Induced Pluripotent Cells
HIV	Human immunodeficiency virus
HMT	Histone methyltransferase
HNRNP	Heterogeneous nuclear ribonucleoprotein
hnRNP	Heterogeneous Ribonucleo Proteins
HP1	Heterochromatin protein 1
HRP	Horseradish peroxidase
HS	Heat Shock
HS	High similarity
HSATIII	Human SATIII repeats
HSE	Heat Shock Elements
HSF	Heat Shock Factor4
HSNR	60` of Heat Shock with No Recovery
HSP	Heat Shock Protein
HSR	Heat Shock Response
HTT	Mutant huntingtin protein
HU	Hydroxy Urea
IC50	Half maximal inhibitory concentration
ICC	Immunocytochemistry
iCLIP	Individual nucleotide resolution cross-linking and immunoprecipitation
IDP	Intrinsic Disordered Proteins
IF	Immunofluorescence

IGF	Insulin-like Growth Factor
IgG	Immunoglobulin G
IP	Immunoprecipitation
IPA	Ingenuity Pathway Analyses
IPTG	Isopropyl-1-thio-β-D-galactopyranoside
ISR	Integrated Stress Response
KAT	Lysine acetyltransferases
KD	Knock-Down
kDA	Kilodalton
KDAC	Lysine de-acetylases
KMT	Lysine Methyltransferase
KO	Knock-Out
LARP	La ribonucleoprotein
LC3	Microtubule-associated proteins 1A/1B light chain 3B
lncRNA	Long Non-Coding RNA
lncSATIII RNA	Long Non-Coding Satellite III RNA
LTR	Long Terminal Repeat
mA	Milliamp
MALAT1	Metastasis associated lung adenocarcinoma transcript 1
MATR3	Matrin 3 Protein
MC3T3	Osteoblast precursor cell line derived from Mus musculus calvaria
MCF7	Michigan Cancer Foundation-7 cell line (Breast Cancer)
MEF	Mouse embryonic fibroblasts
MEM	Modified
MIA	Modular Image Analysis
MMA	Monomethyl Arginine
MMS	Methyl-methanesulfonate
MOV10	Mov10 RISC complex RNA helicase
MPC	Microprocessor Protein Complex
MRN	Mre11/Rad50/Nbs1 complex
mRNA	Messenger RNA
MS	Mass spectrometry
MW	Molecular weight
NCAM1	Neural cell adhesion molecule 1
NCOR	Nuclear receptor co-repressor 1
NEAT1	Nuclear Enriched Abundant Transcript 1
NEDD8	NEDD8 ubiquitin-like modifier
NEDD9	Neural precursor cell expressed, developmentally down-regulated 9
NER	Nucleotide Excision Repair
NF-κB	Receptor activator of nuclear factor kappa B
ng	Nanogram
NHEJ	Non-homologous end-joining
NHS	No Heat Shock
NLS	Nuclear Signal Localisation
NP	Nuclear Protein
NPC	Neural Progenitor Cells
nSBs	Nuclear Stress Body
NTC	Non-Targeting Control

one-way ANOVA	One-way analysis of variance
OPG	Osteoprotegerin
PACT	double-stranded RNA activated protein kinase
PAR	Poly ADP-Ribose
PARG	PAR glycohydrolase
PARP	Poly ADP-Ribose polymerase
PBS	Phosphate buffer saline
PBST	PBS 0.02% tween
PCH	pericentromeric chromatin
PCR	Polymerase Chain Reaction
PD	Pulldown
PDL	Poly-D Lysine
PFA	Paraformaldehyde
PG	Peri chromatin Granules
PK	Protein Kinase
PKMT	Protein Lysine Methyltransferase
pM	picomole
PolyQ	Polyglutamine
PPAR	Peroxisome proliferator-activated receptor
PPI	Protein-Protein Interactions
PRMT	Protein Arginine Methyltransferase
PSA	Prostate-specific antigen
PTM	Post-Translational Modification
PVDF	Polyvinylidene difluoride
R/E	Arginine/glutamic acid-rich motif
RBD	RNA Binding Domain
RBM	RNA Binding Motif
RGG	Arginine-glycine rich motif
RHA	RNA Helicase A
RIPA	Radioimmunoprecipitation assay
RISC	RNA-induced silencing complex
RNAPII	RNA Polymerase II
RPL	Ribosomal protein L
RRM	RNA Recognition Motif
RT-qPCR	Real-time quantitative polymerase chain reaction
S/MAR	Scaffold/Matrix Attachment Regions
SAF	Scaffold Attachment Factor
SAFA	Scaffold attachment factor A
SAFB	Scaffold Attachment Factor B
SAFB1	Scaffold Attachment Factor B1
SAFB2	Scaffold Attachment Factor B2
SAH	S-adenosyl L-homocysteine
SAM68	KH RNA binding domain, containing signal transduction associated 1
SAP	SAF-A/B Acinus and PIAS domain
SATII	Satellite-II
SATIII	Satellite-III
SB	Stress Body
SCA	Spinocerebellar ataxia

SDMA	Symmetric dimethylarginine
SDS	Sodium dodecyl sulphate
SEM	Standard error of the mean
SENP	SUMO specific peptidase 1
SHH	Sonic hedgehog
SH-SY5Y	Human neuroblastoma derived cell line
siRNA	Small interfering RNA
SLTM	SAF (scaffold attachment factor)-like transcription modulator
SR	Serine–arginine-rich
SRRT	Serrate, RNA effector molecule
SRSF	Serine/arginine-rich splicing factor
SSC	Side scatter
STA	Staurosporine
STRING	Search Tool for the Retrieval of Interacting Genes
SUMO	Small ubiquitin-like modifier
TAD	Topologically associating domains
TAE	Tris / acetic acid / EDTA
TAE	Tris-acetate EDTA
TAF15	TATA-binding protein-associated factor 2N
TCP	T-complex protein
TDP	TAR DNA binding protein
TDP-43	Transactive response DNA binding protein (43 kDa)
TEMED	N,N,N',N'-Tetramethyl ethylenediamine
TG	Thioglycerol
TMT	Tandem Mass Tag
TNF	Tumour necrosis factor
TUG1	Taurine upregulated gene 1
Two-way ANOVA	Two-way analysis of variance
TWST	Trainable Weka Segmentation Tool
Ub	Ubiquitin
UBL	Ubiquitin protein ligases
UPR	Unfolded Protein Response
UTR	Untranslated regions
UTR	Untranslated regions
UV	Ultraviolet
VCP	Valosin containing protein
WT	Wild Type
WWP2	WW domain, containing E3 ubiquitin-protein ligase 2
XIST	X-inactive specific transcript
XIST	X-inactive specific transcript
XOR	Xanthine oxidoreductase
XPO1	Exportin 1
µl	Microlitre
µm	Micrometre
µM	Micromole

Chapter 01: Introduction

1.1 Overview of the Scaffold Attachment Factor B Family (SAFB)

The Scaffold Attachment Factor B (SAFB) family of DNA and RNA binding proteins are comprised of three members: SAFB1, SAFB2 and SLTM (SAFB-like transcriptional modulator) (reviewed in [2, 3]). They are large multi-domain proteins that would be predicted to participate in a variety of biological processes. SAFB2 (a paralogue of SAFB1) has 74% similarity with SAFB1 [4], whereas SLTM has only 34% [5]. The same functional domains are expressed in all family members suggesting they have overlapping roles within cell metabolism. Historically, three independent research groups identified SAFB1: Renz and Fackelmayer in 1996, found SAFB1 bound S/MAR (scaffold/matrix attachment regions) in nuclear proteins [6]; Oesterreich et al. in 1997, found SAFB1 bound the Heat Shock 27 (HSP27) promoter as a transcription co-repressor and named it HET (Hsp27 ERE-TATA binding protein) [7]. Later, Weighardt et al., in 1999, identified a protein that interacted with hnRNPA1 and named it HAP (hnRNPA1-associated protein) [1]. SAFB2 was first identified in 2003 by Townson and collaborators as a protein with overlapping functions with SAFB1, such as oestrogen co-repressor, but expressed in the cytoplasm and the nucleus. Additionally, SAFB1 and SAFB2 can interact directly via the C-terminal domain [4]. On the other hand, SLTM was identified as a Mesenchymal Epithelial Transition Factor (MET) by Colley in 2002 [8], and later, it was re-named SAFB-like transcriptional modulator (SLTM) in 2007 [5].

SAFB1 has been studied the most extensively, and it has been shown to modulate chromatin condensation and stabilise heterochromatin foci in animal cell models [9]. It binds H3K9me3 (Histone 3, Lysine 9, tri-methylated) in precipitated chromatin [10] and interacts with several hnRNPs (Heterogeneous Ribonuclear Proteins), and it has been shown to form nuclear granules that co-localise with Heat Shock Factor 1 (HSF1) upon stress exposure [1]. More recently, SAFB1 was shown to be essential for the efficient spreading of γH2AX (phosphorylated Histone 2AX) following a double DNA break [11]. Evidence to date, therefore, suggests that SAFB1, and possibly SAFB2 and SLTM, play roles in chromatin organisation, transcription, splicing and stress response.

1.2 SAFB PROTEINS STRUCTURE

SAFB1 is 917 aa long (accession NP_001188267), SAFB2 932 aa (accession NP_055464) and SLTM 1034 aa (accession NP_079031), which makes them large

proteins that contain similar functional domains. SAFB1, SAFB2 and SLTM have the same 34 base pair SAF-box domain (scaffold attachment factor-box, also known as SAP domain [SAF-A/B Acinus and PIAS domain]), important for DNA binding, a High Similarity (HS) domain of 198 aa containing two SUMOylation recognition sites, an RNA Recognition Motif (RRM) that varies slightly between family members and a Glutamine/Arginine rich domain (E/R-rich domain) responsible for a variety of protein-protein interactions and displaying variation between members. Besides the shared domains, SAFB1 and SAFB2 contain a Nuclear Localisation Signal (NLS) with 17 aa and a Glycine-rich domain (G-rich domain) [2]. The NLS domain in SAFB1 is located before the E/R-rich domain, between aa 599 and 616, whereas SAFB2`s NLS is located within the E/R-rich domain between aa 713 and 730. [2, 3]. SLTM does not have an NLS domain. However, the protein is expressed in both the cytoplasm and nucleus. Figure 1. 1 Domains of SAFB family members. 1) SAF-A/B, Acinus and PIAS (SAP); 2) High Similarity (HS); 3) RNA Recognition Motif (RRM); 4) Nuclear Localisation Signal (NLS); 5) Glutamic Acid/Arginine-rich (E/R-rich domain); 6) Glycine-rich Domain (G-rich Domain). shows the functional domains of the SAFB proteins.

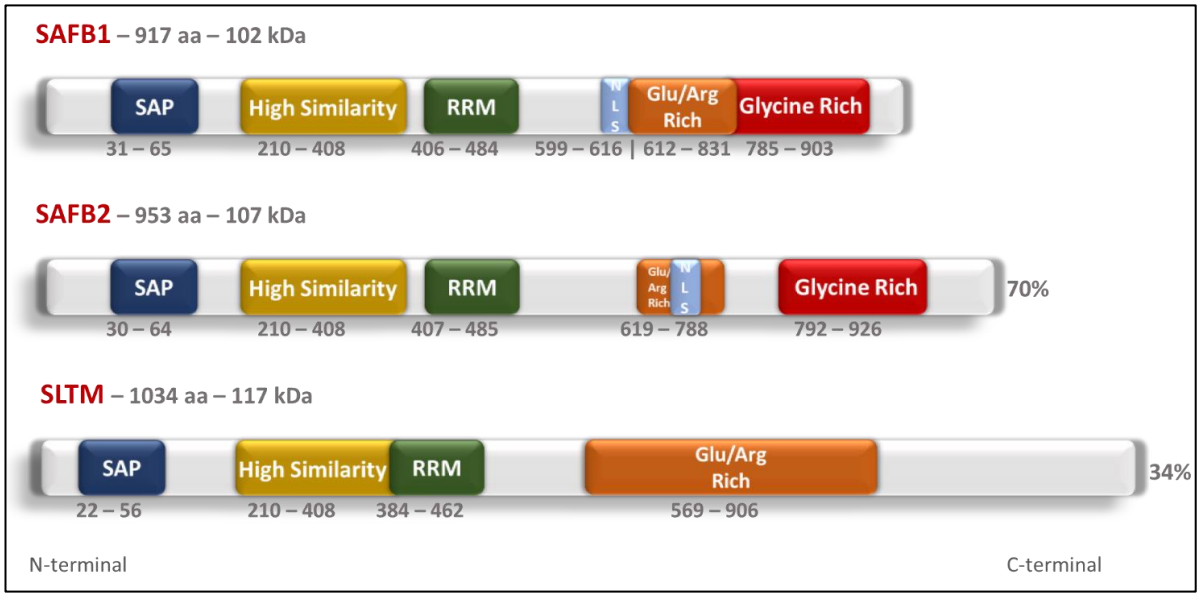

Figure 1. 1 Domains of SAFB family members. 1) SAF-A/B, Acinus and PIAS (SAP); 2) High Similarity (HS); 3) RNA Recognition Motif (RRM); 4) Nuclear Localisation Signal (NLS); 5) Glutamic Acid/Arginine-rich (E/R-rich domain); 6) Glycine-rich Domain (G-rich Domain).

SAP Domain

The N-Terminal SAP domain is homeodomain-like DNA binding motif highly conserved through evolution and present in all family members and it mediates DNA binding. The SAP-box recognises (but is not limited to) A/T-rich motifs in S/MARs, thereby anchoring the double-stranded DNA helix to the nuclear matrix [12, 13]. The SAFB1 SAP domain was found to be required for its localisation to the chromatin-bound fraction and for splicing regulation [14]. The ability to bind to DNA suggests roles in gene expression and chromatin organisation for all the members [3, 9, 15].

1.2.1 High Similarity Domain

The HS is a region of 100% homology between SAFB1 and SAFB2 [2]. Although its function is uncertain, this domain hosts two constitutively SUMOylated Lysines (residues 234 and 291) in SAFB1 and SAFB2 [16], which possibly modulate protein recruitment or interacting partners during cell stages of stress.

1.2.2 RNA Recognition Motif

The RRM is a highly conserved domain present in about two 2% of all human proteins, often in multiple copies [17]. It forms an "αβ-sandwich" structure with four-stranded β-sheets and two α-helices topology (βαββαβ); the two aromatic rings at its central β-sheets are vital for the recognition of RNA [18]. Many proteins involved in the steps of RNA processing, such as synthesis, splicing, transport and degradation, have an RRM [19]. SAFB1 and SAFB2 RRMs` are 98% homologs, while SLTM`s RRM is 70% homolog to SAFB1 [2]. More recently, the RRM domain of all members were found to bind directly to long non-coding SATIII RNA upon stress conditions [20]

1.2.3 Nuclear Localisation Signal

The NLS located at the central region of SAFB1 and SAFB2 is a sequence of positively charged amino acids such as lysine and arginine, responsible for mediating the import of proteins to the nuclear compartment via the nucleopore complex [21]. RNA-biding proteins are frequently observed to shuttle between the nucleus and cytoplasmic compartments to mediate RNA/protein transport [22]. Studies have shown that HeLa cells transfected with an EYFP–SLTM plasmid express the fusion protein in the nucleus but not nucleoli [5]. As SLTM does not contain an NLS, these findings suggest an

association with nuclear proteins [23]. Moreover, this staining pattern partially overlapped with SAFB1 [5], suggesting similar but not identical functions.

1.2.4 Glutamic Acid/Arginine Rich Domain

This important domain is located in the carboxy-terminal region of SAFB proteins. Evidence has shown that the E/R-rich domain can bind protein targets to promote transcription regulation and to modulate cellular stress response [1]. SAFB1 also interacts with Major Satellite RNAs repeat transcripts in animal cell models to promote phase separation of chromatin via E/R-rich domain [9]. Therefore, this region is known to mediate protein-protein interactions. Interestingly, a 73 aa long region between the amino acids 656 and 728 in SAFB1 has 90.5% similarity with SAFB2 and 72% similarity to SLTM (conserved region), suggesting shared E/R-rich domain-dependent functions for SAFB proteins. Additionally, this region is similar to coiled-coiled domains, often seen in homo/heteromeric protein complex formations [24], pointing to the possibility of direct interactions between family members.

1.2.5 Glycine-Rich Domain

The C-terminal G-rich domain is only present in SAFB1 and SAFB2. Pulldown assays performed using the G-rich domains of SAFB1 and SAFB2 have shown that this domain is responsible for direct interaction with Enhancer of Rudimentary Homologue (ERH) [25], a splicing factor required for the expression of multiple cell cycle genes and effective function of ATR (ATM [ataxia-telangiectasia mutated] and Rad3-Related) during the DNA damage response and replication stress [26]. This supports the contention that G-rich domains also mediate protein-protein interactions.

1.3 SAFB Genes / Expression

Mammalian SAFB genes are highly conserved with orthologues among bacteria, fungi, cnidarian, nematodes and arthropods [4], supporting the idea that each has unique functions with the cell metabolism. SAFB1 and SAFB2 genes are paralogues located on the 19p13.3 chromosome, arranged in a head-to-head conformation sharing a CpG dinucleotide rich bidirectional promoter between both sequences [4, 27] (Figure 1. 2 SAFB1 and SAFB2 are on chromosome 19p13, Bi-directionally oriented with a 490 bp promoter located between translational start sites. Image adapted from Townson, S.M., et al, 2003

); whereas the SLTM gene sequence is located on the chromosome 15q22.1 [5].

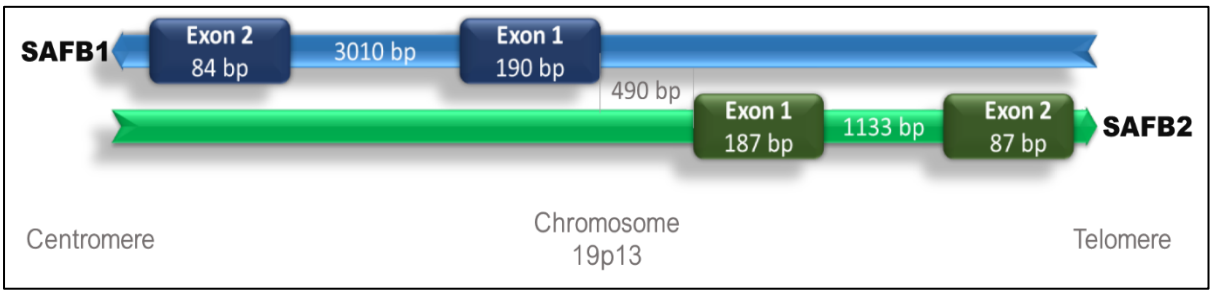

Figure 1. 2 SAFB1 and SAFB2 are on chromosome 19p13, Bi-directionally oriented with a 490 bp promoter located between translational start sites. Image adapted from Townson, S.M., et al, 2003

The SAFB1 genome sequence is 45,443 bp long and encodes twenty-one exons and six validated isoforms ranging between 85 and 102 KDa prior to post-translation modifications (PTMs). Five additional isoforms have been predicted by automated computational analysis using the Gnomon method, according to the Gene database from the National Center for Biotechnology Information Search database (NCBI). Isoform one (accession NM_001201338.2) is recognised as the canonical sequence for SAFB1, with 917 aa and a molecular weight (MW) of 102 kDa prior to PTMs.

Meanwhile, only one SAFB2 isoform (accession NM_014649.3) has been identified with an MW of 107KDa prior to PTMs. However, five isoforms are predicted by computational analysis. Interestingly, Townson et al. suggested that SAFB2 was generated due to a duplication of SAFB1 that underwent an inversion. This also indicates SAFB2 must have unique functions as it was retained via natural selection [4].

SLTM has two validated isoforms with 115.2 KDa and 117.2 KDa prior to PTMs, where the longest transcript variant is the canonical sequence with 1034 aa long. The genome sequence for SLTM is 55264 bp long, encoding for twenty-three exons (accession NM_024755.4). Moreover, it has two validated transcript variants of non-coding RNA (ncRNA) [28], and sixteen isoforms are predicted by computational analysis [29].

The SAFB1 protein is expressed across all tissue types in the human body, but it has high expression in the Central Nervous System (CNS), especially in the cerebellum [30]. Other tissues with high SAFB1 expression are the thymus/immune system [31], prostate [32] and placenta [33] in adult life. Decreased levels of SAFB1 were detected during the differentiation of the 3T3-L1 cells into adipocytes and Caco-2 mediated differentiation into enterocytes [34], whereas it is expressed and important for embryonic development and early-life growth [35].

Similarly, SAFB2 is expressed in all human tissues as well, with higher expression in the immune system and CNS, although its expression is not as high as SAFB1 [36]. SAFB2 is also expressed in the reproductive system during the early and late stages of life [31]. SLTM, on the other hand, is more expressed in Bone Marrow, Thyroid, lymph nodes and fat [36]. Interestingly, SLTM expression is necessary during the intra-uterus development/growth, as its absence causes perinatal lethality [37].

In summary, both SAFB1 and SAFB2 proteins contain the NLS, which directs them to the nucleus after protein synthesis [38], where SAFB1 associates with heterochromatin, playing a role in chromatin organisation and condensation [9]. SLTM, on the other hand, lacks the NLS, although literature reports have shown SLTM in co-localisation with SAFB1 within the nucleus compartment, except for the nucleoli. SAFB1 distribution is nuclear, offering small and diffuse granules throughout the compartment, excluding nucleoli [1]. SAFB2 shows a similar pattern, but it is expressed in the cytoplasm as well. Moreover, the staining of SAFB2 with specific antibodies for fluorescence microscopy displays a lower intensity in the cytoplasm than the nuclear staining [4]. Meanwhile, SLTM staining follows the same trend as SAFB1 despite its lack of NLS domain and co-localises with SAFB1 in the nucleus. [5].

1.4 FUNCTIONS OF SAFB PROTEINS

The presence of multiple domains in SAFB proteins allows family members to interact with many classes of biomolecules within cells. SAFB1 has been shown to interact with DNA [6], RNA [30] [39], transcriptional factors [25] and hormone receptors [7, 40] amongst other important proteins, allowing functions on different processes and pathways, such as chromatin organisation/condensation [6, 9, 15, 41], gene expression [4, 5, 34, 42-50],

splicing [14, 16, 30, 49], stress response [1, 15, 27, 41] and DNA Damage [11], whereas SAFB1 and SAFB2 have been recently shown to bind the microprocessor complex (MPC), as accessory proteins [51]. More recently, all 3 family members were shown to co-immunoprecipitate with an anti-sense SATIII RNA probe in a comprehensive identification of RNA-binding proteins by mass spectrometry (ChIRP-MS) study [20]. Interestingly, SLTM and SAFB2 held the 1st and the 8th highest scores, whereas SAFB1 occupies the 206th place on the list of lncSATIII binding partners.

Past studies have proven the importance of SAFB proteins in mice embryos' viability and/or development [31, 35, 37]. Knockout experiments revealed SAFB1 to be vital for healthy growth and development as SAFB1-/- homozygous mutant mice showed high prenatal/neonatal lethality with Intra-utero growth delay. In a further development, the viable animals displayed low serum levels of Insulin-like Growth Factor (IGF), leading to dwarfism and defects in the reproductive system [35], which agrees with SAFB1 role as a co-repressor for hormone receptors [34, 40]. Another knockout (KO) study showed that SAFB1-null mice developed abnormalities in their lungs and different types of tumours throughout growth [52]. Moreover, several defects on the immune system of the SAFB1 depleted mice were found, showing the importance of SAFB1 for the healthy function of mammalians across different stages of life. Deleting the SAFB1 gene sequence in mouse embryonic fibroblasts (MEFs) increased the number of double-stranded breaks (DSBs) compared to wild type MEFs. Such an increase in DSBs could lead to mutations that cause either the inactivation of tumour suppressor genes or activation of oncogenes. This alteration caused spontaneous immortalisation and altered expression of two other proteins involved in immortalisation and escape from senescence: low levels of CDKN2A and high levels of TBX2 [42]. Therefore, the evidence shows that SAFB1 is important for the efficient function of several pathways and mechanisms with cell metabolism.

SAFB2$^{-/-}$ homozygous mutant mice did not display obvious defects in growth or fertility, although later analysis pointed to an increase in testis weight and Sertoli cell count [31]. The authors suggested this could be caused partially due to defects in androgen-receptor function and expression in the SAFB2 KO mice. Therefore, despite their similarity, evidence suggests SAFB1 and SAFB2 have distinct functions.

Finally, SLTM $^{-/-}$ knockout mice generated by CRISPR/Cas9-mediated gene editing led to perinatal lethality without any other apparent defects [37]. Moreover, SLTM $^{-/-}$ MEFs showed higher expression of Gil1 and Ptch1; when overexpressed, SLTM leads to repression of these genes, pointing to SLTM as a repressor of SHH/GLI genes.

The mapping of Protein-Protein Interactions (PPI) allows us to understand the pathways involved in biological pathways and how these pathways promote the efficient functioning of an organism [53]. Identification of protein hubs – proteins exhibiting large degrees of interactions with other proteins – might be the key to understand pathways crosstalk and how the miss-functioning of proteins can influence several different biological functions. These protein hubs challenge the current Protein Structure–Function Paradigm, which state that to exert its functions properly, the amino acid sequence of a protein must be arranged in a three-dimension (3D) structure that would act as a lock-and-key mechanism to promote interactions with limited specific proteins. However, not all proteins have a defined 3D structure, named ordered proteins, and a higher number of important proteins are found not to fit this model [54]. Therefore, in 2001, Dunker and collaborators proposed that proteins could exist in three different states: ordered state, molten globule and random coil [55]. Intrinsically Disordered Proteins (IDP) would then be proteins that exert their functions differently from ordered. IDPs have many interaction partners due to the lack of a frozen 3D structure and a certain level of adaptability, called the promiscuity level.

Interestingly, the analysis of human hubs with a higher-than-expected PPI density in MCF7 cells shows that SAFB1 has a high Intrinsic Disorder Interactions (IDI) score and is one of the most disordered proteins with elevated promiscuity levels [53]. In this study, SAFB1 was found to host 18 Anchor-Identified biding sites for protein-protein interactions, 818 chromatin regions were found to be enriched in SAFB (SAFB1/SAFB2) ChIP experiments, where 541 of them were placed within promoters of known genes, whereas 277 chromatin enriched regions were upstream of unknown genes. SAFB biding clusters were identified in chromosomes one and six, especially within histone gene clusters, reinforcing the theory of SAFB functions in a higher-order chromatin organisation. SAFB1 specific gene targets were related to immune regulatory genes (38%) and signalling related genes (28%) [43]. The study also showed that proteins associated with high-

density disorder-enriched hub PPIs are characterised by similar evolutionary conservation levels [53], which is in accordance with the evolutionary conservation of SAFB proteins [4].

In a neuro-disease panel, α-synuclein, [56], amyloid β and tau protein, prions, ataxin [57], as well as TAR-DNA binding protein-43 (TDP-43) and the RNA-binding protein FUS [58] were found enriched within the human protein hub as IDPs. From a cancer point of view, p53 [59] and BRCA1 [60] are also IDPs found with a higher promiscuity degree [53]. Interestingly, SAFB1 has been proven to associate with TDP-43 and FUS [14, 61], as well as with p53 [50] and BRCA1 [62], which suggests that SAFB1 could be somehow involved in these conditions.

1.4.1 SAFB in Chromatin Organisation

Chromatin is organised hierarchically in different architectural patterns, such as compartments, chromatin loops and topologically associating domains (TADs), which confers to the genome the necessary proximity between regions that are not directly linked but still influences their fate [63]. For instance, the zinc finger CRCF (CCCTC-binding factor), which has a dual-functionality as a gene activator/repressor and as an insulator, was enriched within the boundaries of TADs and is required for the formation of loop domains [64]. Other patterns found within the limits of topological domains were housekeeping genes, tRNAs and retrotransposons [65]. Liquid-Liquid phase separation is another mechanism proposed to explain the dynamicity of chromatin regions [66]. All of these play a role in chromatin organisation and in maintaining the chromatin hierarchy.

It is established that SAFB1 has functions in chromatin organisation. However, the mechanisms and preferences that rule SAFB1-DNA biding and its function are still poorly understood. Recently, a study from Huo and collaborators [9] had shown that SAFB1 is required for pericentromeric chromatin (PCH) stabilisation, is associated with different RNAs, including highly repetitive transcripts such as MajoSAT RNAs in mice (equivalent to SATIII transcripts in human cells) and drives phase separation promoted by the MajoSAT RNAs, therefore regulating Chromatin condensation via remodelling higher-order of PCH. Additionally, SAFB depletion significantly decreases H3K9Me3 foci formation, a marker of chromatin condensation and gene silencing [reviewed in [67]],

reinforcing SAFB1`s role as a transcription repressor. Overexpression of an EGFP-tagged SAFB construct is found to have a subnuclear localisation close to DAPI dense regions, known to correspond to mouse PCH. The EGFP–tagged SAFB1 co-localises with H3K9Me3 regions and HP1α foci [9]. HP1α foci bind with H3K9Me3, forming a heterochromatin packing marker especially present in centromeres and telomeres of Eukaryotic cells [67]. Hi-content 3D images showed EGFP-tagged SAFB recruitment to the periphery of PCH, whereas the deletion of SAFB reduced not only H3K9Me3 but also SUV39H1 and H4K20me3, both heterochromatin markers [9]. Moreover, distances between HP1α and DAPI-dense PCH were increased in SAFB depleted cells, suggesting a disruption of chromatin structure.

In agreement with these results, Ninomiya and collaborators found that SAFB1, SAFB2 and SLTM interact with lncSATIII repeats via their RRM domains following a thermal stress [20]. Although the authors suggested a role in splicing following a stress, the results suggest SAFB proteins play a role in tethering RNA and proteins together in the nuclear stress bodies. Taken together, these results point to a role for SAFB proteins in chromatin organisation and high architectural order.

1.4.2 SAFB proteins in transcription and RNA Processing

SAFB1 has been proven to bind the HSP27 promoter, and its overexpression represses the HSP27 promoter activity in a breast cancer cell line in a dose-dependent manner [7]. Further publication from the same group showed that the C-terminal region mediated the repressive role of SAFB1. This region was also found in SAFB2, suggesting a repressive role for SAFB2 [47]. This hypothesis was later proven to be true by Hashimoto and collaborators in a paper in which they reinforced the repressive role of the proteins and showed that SAFB1 and SAFB2 decrease oestrogen Receptor alpha (ERα) mobility [68]. Moreover, p53 transcriptional activity is also suppressed by SAFB1 via C-terminal interactions [50].

In accordance, SAFB1 was further presented as a general co-repressor for nuclear receptors, such as FXRα (Farnesoid X Receptor alpha), PPARγ (peroxisome proliferator-activated receptor-gamma) and Gal4-PPARγD/E. This activity was dependent on SAFB1 C-terminal domains. [34]. The authors of this study hypothesised that SAFB1 could

potentially influence proliferation and differentiation decisions during development. The Influence of SAFB1 in cell proliferation was validated in 2006 by Dobrzycka through knockout experiments, which led to increased proliferation rates in KO MEFs cells [42].

Since many studies have demonstrated a role for SAFB1 in transcription, the next step would be identifying an RNA target sequence consensus. In 2015, our lab used individual-nucleotide resolution Cross-Linking and Immuno-Precipitation (iCLIP) and showed that SAFB1 binds GAA/AAG/AGA motifs in exons [30]. The Cell Adhesion Molecule (NCAM1) and Astrotactin 2 (ASTN2) mRNAs were shown to be targeted and modified by SAFB1, and two of three NCAM1 spliced forms were altered by SAFB1 depletion. The same study showed SAFB1 bound several long non-coding RNAs (lncRNA) such as MALAT1, NEAT1 and TUG1, all of them rich in GAA/AAG/AGA motifs, [30]. Another study used Chromatin Immunoprecipitation (ChIP) and ChIP-on-chip assays to identify 818 SAFB1/SAFB2 target sites on chromatin, of which 541 were found to map to promoters of known genes [43]. Downregulation of SAFB proteins increased the expression of 457 genes, supporting its repressive transcription role, and decreased the expression of another 259 genes, suggesting it acts as an activator. Among all 716 genes identified, SAFB1 regulated 680, 163 of them were not affected by SAFB2 depletion, and 24 promoter gene sequences were found to be targets of both SAFB1 and SAFB2 proteins [43]. Another important finding is that the SAFB1 gene repression role coincides with histone methylation. SAFB1 co-immunoprecipitates with EZH2 (enhancer of zeste 2 polycomb repressive complex 2 subunits), the major histone methyltransferase, which catalyses the trimethylation of histone 3 Lysine 27 (H3K27me3), a known marker for gene silencing [69, 70]; depletion of SAFB1 reduced the accumulation of H3K27me3 on the prostate-specific antigen (PSA) promoter significantly [31], suggesting a role for SAFB1 in cancer.

Additionally, Fused-in-Sarcoma (FUS), a nuclei-cytoplasmic protein with roles in RNA regulation and metabolism [71], requires the SAFB1 SAP domain for its localisation to chromatin to exert its roles in splicing regulation [14]. Interestingly, a mutant version of FUS present in ectopic ALS (Amyotrophic Lateral Sclerosis) can sequester Matrin3 and SAFB1 in cytoplasmic aggregates, linking SAFB1 to ALS. A more recent study showed SAFB2 as part of alternative RNA splicing complexes, acting as a negative regulator of a tra2β variable exon [72]. The same research used gradient fractionated HEK293 nuclear

extracts to show that SAFB2 is present in the nucleus compartment associated with high molecular mass protein complexes (up to 670 kDa), whereas SAFB1 is present as a monomer or in small nuclear protein complexes. Moreover, overexpression of SLTM has also been shown to act as a generalised transcription inhibitor [5].

In addition to SAFB roles in transcription, SAFB1 was found within spliceosomes [73] and our group showed its knockdown reduced the processing of miR-19A from the miR-92 cluster (miR-17,18a,19a,20a,19b,92) [30]. Hutter and collaborators in 2020, showed SAFB1and 2 bind to the microprocessor complex, and it is essential for efficient processing of Pri-miR-15a and several clustered miRNAs [51].

1.4.3 SAFB proteins in cellular stress response/DNA Damage and Repair

Possible involvement of SAFB proteins in the cellular stress response was first reported in 1999 by Biamonti's research group. Weighardt et al. demonstrated that SAFB, termed HAP at the time, formed aggregates within the nucleus following heat stress [1]. Specific aspects of the Heat Stress Response (HSR) will be discussed in detail in the following sections. These HAP bodies also overlapped with HSF1 stress bodies, and their formation was dependent on RNA synthesis. Further studies investigated the chromosome localisation of heat stress-induced HSF1 and hnRNP A1-interacting protein aggregates, such as SAFB proteins (termed HAP at the time) [74]. The authors demonstrated that HAP aggregates do not occur in hamster cells and used hamster:human cell hybrids to precisely map the stress bodies to chromosome. The results showed HAP stress bodies' formed on the pericentromeric heterochromatic q12 band of chromosome nine and secondary assembling on centromeric regions of chromosomes 12 and 15. Moreover, the 9q12 region is rich in Satellite DNA (SATIII). SATIII DNA does not encode for protein but is transcriptionally active upon heat stress, producing a long non-coding RNA (lncRNA) – named lncSATIII [75]. In addition, Chromosomes 12 and 15 were able to form SAFB1 stress bodies in co-localisation with an αSAT probe. Alpha satellite DNA (αSAT) is a tandemly organised type of repetitive DNA found at all human centromeres [76], locating secondary HAP aggregates to the centromeric region of these chromosomes. However, it is important to point out that there is no further information about the specificity of the HAP antibodies (see below). More recently, ChIRP-MS experiments using lncSATIII probes as bait have shown that SLTM, SAFB2, and to a lesser degree

SAFB1, interact directly via RRM domains with the newly synthetized stress induced lncSATIII transcripts [20].

Two further studies of SAFB1 protein function were published using the HAP-antibody, which is predicted to recognise SAFB1 and SAFB2 [77, 78]. Chiodi et al. showed that SRSF1 (Serine-Arginine splicing factor 1) is recruited to the HAP aggregates upon Heat shock exposure via the second RRM domain (RRM2) present in the SRSF1 [77]. In contrast, Valgardottir applied several stress inducers, such as cadmium sulphate, hyperosmotic stress, UV radiation, and Etoposide, to test whether they could form HAP aggregates and induce the transcription of SATIII [78]. Cadmium sulphate and hyperosmotic stress drove the transcription of SATIII, and the assembling of HAP aggregates in similar levels of heat stress, whereas UV radiation and Etoposide treatment induced the stress response in a limited percentage of the cells. All treatments that demonstrated an increase in the transcription of SATIII also found HAP aggregates were formed [78]. The authors hypothesised that the formation of HAP aggregates and upregulation of SATIII is somehow related.

Currently, there are limited reports on SAFB2 involvement in the heat stress response. However, due to the presence of many shared domains between SAFB1 and SAFB2, they might have several overlapping functions within cell metabolism. Significantly, the majority of the above-mentioned studies [1, 74, 77, 78] used a polyclonal antiserum affinity-purified on recombinant His-tagged HAP as an anti-HAP (SAFB) antibody. However, due to the age of the papers and the antibody production technique, no information could be found regarding antibody specificity. However, the HAP-antibody was raised against regions of high similarity between SAFB1 and SAFB2, and the HAP antibody is likely to recognise both proteins.

SAFB1 also plays a role in the DNA Damage Response. SAFB1 depletion leads to the impairment of the spreading of phosphorylation of Histone 2AX foci (ɣH2AX) following DNA damage induction by ionising radiation (IR) in U-O-2 cells [11]. Hence, SAFB1-depleted cells were less efficient in triggering the G2/M checkpoint and showed reduced survival in response to the replication fork stalling inducer hydroxyurea (HU). Furthermore, the reduction of ɣH2AX signal was not a result of fewer DSBs, but due to a decrease in the capacity of the DDR machinery to process efficient signalling following

DNA damage. The study also performed a drug screen to find compounds capable of modulating the TopBP1 ATR activation-construct-induced system in a DNA-damage- and replication-independent γH2AX manner. Interestingly, the twelve top-scoring compounds found were all inhibitors of histone deacetylases (HDACs), suggesting that SAFB1-mediated enhancement of γH2AX signalling cooperates with histone acetylation to overcome chromatin constraints that limit DDR signalling. Therefore, SAFB1 collaborates with histone modification to efficiently spread γH2AX, a marker of DNA damage, and is necessary to activate DNA repair machinery. Laser micro irradiation experiments performed by the same group also showed that SAFB1 is excluded from DNA damage sites minutes after damage induction. However, after combined mild detergent-mediated pre-extraction with RNase digestion to remove the soluble and RNA-bound protein pool, the remaining SAFB1 was transiently recruited to sites of laser micro irradiation followed by persistent release from the damaged area. This means that the SAFB1 fraction not involved in RNA processing accumulates in the DNA damaged surrounding areas. Additionally, the authors showed that co-depletion of PARP1 and PARP2 (poly-ADP-ribose polymerase 1 and 2) resulted in the abolishment of SAFB1 recruitment from the DNA damage surrounding. In contrast, the depletion of the antagonist of PARP-dependent chromatin PARylation, PAR glycohydrolase (PARG), enhanced SAFB1 recruitment, suggesting that PARP1 is mainly responsible for the observed SAFB1 recruitment. Analysis of SAFB1 deletion-mutants revealed that the SAFB1 R/G-rich domain, specifically the motif RGGMSGRG containing one RGG and one RG site (R868, R874) within the c-terminal portion, was required and sufficient for SAFB1 recruitment, which suggests that this region mediates its PARP1- and PAR dependent accumulation at damaged chromatin. On the other hand, enforced expression of SAFB1 at the sites of DNA damage increased the spreading of γH2AX [11], reinforcing the hypothesis that SAFB1 recruitment may assist the early chromatin response to DNA breakage. However, its subsequent exclusion might help prevent excessive signalling.

Immunoprecipitation assays demonstrated an interaction between SAFB1 and Ku86 [79], which is an essential component of the no-homologous end-joining (NHEJ) machinery in mammals [80]. SAFB1 co-immunoprecipitates together with SAFB-A and BRG1 (chromatin-associated protein) as part of a cluster of transcriptional proteins [79], which represses the Xanthine oxidoreductase gene (hXOR). SAFB1 binds the hXOR promoter

and restricts its expression via interaction with E-box, DNA-PKs and tumour suppressors [81]. hXOR is an important source of reactive oxygen species implicated in the immune response, and its miss-expression is associated with human diseases, including cancer. The study used a Rabbit Anti-SAFB1 antibody, provided by Beth laboratories and no information on antibody specificity was found.

1.4.4 SAFB proteins in apoptosis

Our group showed SAFB1 plays a role in apoptosis induced by staurosporine (STA), camptothecin and etoposide in HeLa, SH-SY5Y and COS cells [82]. SAFB1 was cleaved into several fragments 30 minutes after STA induced apoptosis and the treatment re-distributed SAFB1 to the nucleoli, where it is normally excluded. Longer exposure to STA caused the formation of a transient peri-nucleolar ring of SAFB1, no longer visible three to four hours following treatment. The dissolution of SAFB1 rings overlapped with chromatin condensation, which again agrees with the role of SAFB1 in chromatin organisation [9]. More importantly, the SAFB1 peri nucleolar ring did not co-localise with HSF1. Moreover, the peri nucleolar rings were sensitive to RNAse treatment, but not DNase, pointing to a role of RNA transcripts in SAFB1 function within the apoptotic process. The same study demonstrated that the recruitment of SAFB1 was through the E/R-rich domain. Thereby, the authors suggested that SAFB1 has an important role in the apoptotic process.

Chromatin immunoprecipitation using an in-house made pan SAFB antibody following SAFB1 or SAFB2 KD revealed that 11% of SAFB targets are involved in apoptosis, whereas loss of SAFB1 de-repress oestrogen-regulated genes, such as BBC3, NEDD9 and OPG [43]. BBC3 (also known as PUMA - p53 upregulated modulator of apoptosis) belongs to a pro-apoptotic subclass of proteins that induce mitochondrial dysfunction [83], whereas NEED9 (neural precursor cell expressed, developmentally down-regulated 9) downregulation induces apoptosis and cell growth [84], and OPG (Osteoprotegerin) induces apoptosis of osteoclasts osteoclast precursor cells partly by the Fas/FasL signalling pathway [85].

Furthermore, SAFB2 has not yet been implicated in the apoptotic process. However, studies involving overexpression of SLTM in HeLa, MCF-7, MC3T3 cells, HepG2 cell lines

and human fibroblasts have demonstrated an increase in the percentages of cell populations at the pre-G1 cell cycle phase. Microscopy analysis of these cells showed an increase in the apoptotic index and cells displaying DNA condensation and nuclear fragmentation, both compatible with cells undergoing an apoptotic process. Moreover, cells overexpressing SLTM also showed signs of extensive Cytochrome C release [5].

1.4.5 SAFB in Cancer and Disease

The BRCA1 gene (breast cancer susceptibility gene 1) is a tumour suppressor that acts as an ERα corepressor [86], similar to SAFB1 and SAFB2. BRCA1 has been involved in 5-10% of all cases of breast cancer (50% of all heritable cases) [87]. Due to SAFB1/SAFB2 and BRCA1 function resemblance as an ERα corepressor, it has been proposed that SAFB proteins could be involved in cancer as well [88]. Indeed, in 2001 Oesterreich and collaborators found a high frequency of loss of heterozygosis near the SAFB locus. Additionally, low levels of SAFB were associated with a worse prognosis for breast cancer patients [89, 90]. Interestingly, overexpression of RNA binding proteins leads to apoptosis in cancer cells [91], corroborating with recent studies where overexpression of SAFB1 inhibited cell proliferation in vitro and decreased the number of cells in the S-phase [92], revealing a possible gene target for future anti-cancer therapies.

Furthermore, lower SAFB1 mRNA levels were found in prostate cancer cells, especially in metastasised cells, compared to normal prostate tissue [32]. The study authors demonstrated that loss of SAFB1 increased cell growth in culture and elevated the levels of Androgen-Receptor (AR) and Prostate Specific Antigen (PSA). As a co-repressor of transcription, low levels of SAFB1 leads to the upregulation of many important genes, such as cell signalling genes [43]. Therefore, these characteristics might be related to the genesis/progression of cancer. As mentioned before, deletion of SAFB1 in MEFs causes cell immortalisation, cell transformation and loss of contact growth inhibition [42], reinforcing the idea that SAFB1 has a key role in preventing loss of cell growth.

SAFB proteins have been implicated in several other diseases. For instance, Takeuchi et al. (2017) [93] has identified autoantibodies against SAFB proteins in patients with Interstitial Lung Disease. Another study published in 2017 [94] has shown that SAFB1 represses HIV-1 transcription by binding to the 5` LTR (long terminal repeat) promoter of

the HIV-1 and successfully impairs transcription initiation and elongation of the virus. SAFB1 inhibited the recruitment of phosphorylated RNA Polymerase II to the 5` LTR, a process that was mediated via the SAFB1 C-terminal domain, specifically the Gly-rich domain, suppressing HIV-1 infection in $CD4^+$ cells. Moreover, the presence of multiple targets of SAFB1 in the NCAM1, ASTN and PDE4B genes is known to play important roles in regulating synaptic function and are implicated in human psychiatric disease. Together with the high level of SAFB1 constitutive expression in the hippocampus [29], these findings suggest that SAFB1 regulates the expression of genes important for synaptic integrity and memory formation.

Our group has recently reported SAFB proteins are linked to the aetiology of the spinocerebellar ataxias (SCA) and Huntington's disease (HD) and may be a general marker of polyglutamine repeat expansion diseases [95]. Abnormal expression of SAFB1, with increased expression in the nucleus and extensive expression in the cytoplasm where SAFB1 co-localized with markers of Purkinje cell injury were found in *postmortem* brain tissue of SCA patients. In HD patients, SAFB1 expression was increased in the nucleus and cytoplasm of striatal neurons. Further investigation found increased binding of SAFB1 to pathogenic ATXN1-85Q mRNA [95].

There is considerable evidence linking the aetiology of polyQ disorders to the altered function of miRNAs, Ago2 and Pum1 [96-101]. The expression of ATXN1 and ATXN3 was regulated by specific miRNAs whose altered expression increased disease pathology [102, 103]. The RBP (and Ago2 binding protein) PUM1 was recently shown to regulate ATXN1 expression directly, and its downregulation increased neurodegeneration [99]. Expression of mutant HTT also leads to a Dicer-mediated increase in the levels of small CAG-repeated RNAs (sCAGs) that caused neuronal death via an Ago2 dependent mechanism [98]. Furthermore, HTT has also been shown to be involved in RNA transport in neuronal dendrites where it associates with Ago2 to repress translation [96, 97]. Importantly, aggregation of mutant HTT (mHTT) was shown to impair LC3 mediated autophagy leading to Ago2 accumulation and global alterations in miRNA levels and activity [100].

SAFB, stress bodies and relevance to disease

In response to proteotoxic stress, constitutive transcription and translation are halted, and the synthesis of protective molecular chaperones such as HSP70 (HSPA1A) is mediated [1]. HSF1 is activated following cellular stress (being released from an inhibitory, HSP90/HSP70/HSP40 complex). This drives the expression of molecular chaperones that protect nascent proteins, refold damaged proteins and help remove potentially harmful proteins and aggregates [1]. HSF1 also drives long non-coding satellite III repeat RNA expression that forms the matrix for nuclear stress bodies (nSB). nSBs bind splicing/transcription factors (SAFB1, SRSF1/9, Sam68) and their temporary sequestration allow gene expression to be suppressed during stress and rapidly recover after stress [20, 104-107]. I have found found that SAFB2 is also a major constituent of nSBs and another member of our group showed that SAFB1 regulates HSP70 (HSPA1A) expression. We also found that SAFB1/2 also binds HSP90 (HSP90AA1 and HSP90AB1) and that this binding is significantly increased ($P<0.001$) following cellular stress. Furthermore, transcriptomic analyses following SAFB2 knockdown showed a significant reduction ($P<0.002$) in HSP90 expression (GEO Series accession number GSE75469) [30]. Studies show cells expressing polyglutamine expansions may not be able to cope with the aggregates generated due to: (i) an age-related reduction in the capacity of protein degradation pathways; (ii) the high level of aggregate production; (iii) an inability to mount a stress response, possibly due to an altered chromatin architecture interfering with HSF1 mediated transcription [108, 109]. SAFB proteins are known to modify chromatin structure [3], and we have shown that SAFB1 is needed for HSP70 (HSPA1A) to be maximally induced following stress. In addition, HSP90 and HSP70 regulate the activation of HSF1. Together these data suggest the abnormal expression of SAFB1 seen in SCA and HD (discussed in section??) will reduce the ability of neurones to mount an effective stress response and alter the splicing and expression of coding and non-coding genes.

1.5 Post-translational modification of SAFB proteins

Recent studies using Immunoprecipitation techniques coupled with mass spectrometry of Hela, HEK293, ESCs (Embryonic Stem Cells) and MV4-11 (human acute myeloid

leukaemia cell line) have shown that SAFB proteins are modified by acetylation [110, 111], phosphorylation [112-115], SUMOylation [44, 116-121] and methylation [122].

A common co- and post-translational modification in cell metabolism is acetylation, which involves adding an acetyl group to the N-terminal end of histone and non-histone proteins catalysed by n-terminal acetyltransferases (NATs) [123]. Lysine's are the most targeted residues where acetylation occurs, and the reaction is catalysed by Lysine acetyltransferases (KATs) [110]. Lysine acetylation generated by KATs is reversible and neutralises lysine`s positive charge, whereas de-acetylation reaction is catalysed by Lysine de-acetylases (KDACs) [124]. Acetylation and deacetylation of histone proteins have a profound effect on gene expression, acting in part to 'open up' chromatin for appropriate transcriptional machinery to access the DNA template [125], and these reactions are carried out by histone acetyltransferases (HATs) or histone deacetylase (HDACs), respectively [110]. SAFB1 are acetylated at Alanine2 (A2) and Lysine607 (K607), whereas SAFB2 is acetylated at residues A2 and K6016. Interestingly, SLTM, have 3 acetylated residues: A2, K401 and K1024 [110, 111].

The phosphorylation/dephosphorylation of a protein is a dynamic, reversible event, dependent on the activity of both kinases and phosphatases, where a phosphate radical is added or removed from a protein [126]. It provides a rapid signal that quickly modifies protein-protein interactions and is used by cells to respond to stimuli. Moreover, it is the most common form of PTMs and is used in most signal transduction pathways. Renz and Frackelmayer first demonstrated the phosphorylation of SAFB [6]. The authors carried out immunoprecipitation of SAFB proteins from nuclear extracts of HeLa cells labelled with 32^{P} in vitro to prove that SAFB has multiple phosphorylation sites. Later on, Nyler and collaborators purified overexpressed mouse CDC2-like kinase (mCLK2), a member of the Serine/Arginine protein kinases, from A293 cell lysates using anti-mCLK2 antibodies and used this in an in vitro kinase assay with the purified recombinant p43 fragment of human SAFB (hSAFB) as a substrate. Phosphorylation of hSAFB by mCLK2 was shown to happen, and the authors concluded that mCLK2 is a potential kinase for SAFB and can thereby regulate SAFB biological activity [127]. Subsequently, studies have shown SAFB1 to be phosphorylated at several residues (Serine's 24, 55, 79, 195, 197, 209, 383, 415, 582, 601, 604 and at Threonine 195), either in response to stimuli,

growth or stress, transiently or constitutively [112-115, 128, 129]. SAFB2 is also phosphorylated at similar residues (Serines 54, 109, 158, 207, 507, 513, 832, 886 and Threonine 201) [113-115, 128-130], whereas SLTM is phosphorylated at Serines 93, 97, 139, 144, 289, 421, 550, 551, 553, 748, 789, 800, 815, 909, 923, 929, 944, 998, 1002, 1014, 1019 and 1021 [113-115, 128, 131, 132].

SUMO is a small ubiquitin-like modifier comprised of 101 aa that shares 20% of similarity with ubiquitin and can be bound covalently to proteins by the SUMO-E3 ligases enzymes through a post-translational process called SUMOylation [133]. Targeted motifs for SUMO addition are a four amino-acid sequence of ΨKXE, where Ψ is a hydrophobic amino-acid, K is the target Lysine, X is any amino-acid, and E is a Glutamic Acid [134]. De-SUMOylation is the reverse process carried out by the SEMP family of proteins [135]. Quantitative proteomic approaches have shown upregulation of SUMOylation in proteins involved in double-strand break repair, methyltransferase activity and signal transduction, all in response to methyl-methanesulfonate (MMS), an alkylating agent that induces replication fork progression impairment via the formation of DNA adducts [136]. The same study also found downregulation of SUMOylation in proteins involved in transcription, transcription co-factors and chromatin modifications and organisation following MMS treatment. Overall, the majority of SUMO changes were enriched in the nucleus and chromatin, suggesting that SUMOylation is a dynamic process involved in the stress response and transcription processes.

A study conducted by the Oesterreich group showed that SAFB1 is SUMOylated by SUMO1 and SUMO2/3 at lysine's K231 and K294, and SEMP1 is the enzyme responsible for de-SUMOylating SAFB1 in MCF7 and HEK293T cells, whereas, in mouse testis cells, SAFB1 is SUMOylated at K316 and K373 [44, 137]. SAFB1 SUMOylation impairment was shown to decrease SAFB1's interaction significantly with HDAC3 (Histone deacetylase 3), a known transcriptional repressor and SAFB1 binding partner, leading to the partial loss of SAFB1 repression activity [44]. Another study proved that SUMOylated SAFB1 prompts the binding of RNA polymerase II to gene promoters and stimulates pre-mRNA splicing [16]. Therefore, mutations on SAFB1 SUMOylation sites prevent the protein from exerting its repressive activities. Subsequently papers have demonstrated other SAFB1 SUMOylation sites at Lysine's 172, 186, 231, 294, 381, 392, 483, 514, 543,

570, 578 and 847 [44, 116, 120, 121], whereas SAFB2 is SUMOylated at Lysine's 65, 94, 188, 199, 230, 293, 380, 385, 388, 391, 395, 517, 524, 525, 541, 542, 551, 578, 586, 608 and 616 [44, 116, 117, 120, 121, 138]. Additionally, SLTM has been proven to be SUMOylated at Lysine's 500, 884, 1024, 1027 [116, 120, 121].

SAFB2 was also found to be ubiquitinated by the BRCA1/BARD1 heterodimer, which exhibits a ubiquitin ligase activity [62]. SAFB2 ubiquitination led to increased expression levels of the protein, whereas overexpression of SAFB2 reduced the BARD1 expression, but not BRCA1, evidencing a possible feedback mechanism between SAFB2 and BARD1. Since BRCA1 is a key protein involved in breast cancer, there might be a link between SAFB2 and tumorigenesis.

Methylation is a post-translational modification of histone and non-histone proteins that adds one, two or three methyl groups to either arginine's or lysine's residues. Histone proteins were the first discovered class of methylated proteins [139], which render chromatin-dependent transcriptional repression or activation [140]. However, histone proteins are not the only targets for methylation and methylation of non-histone proteins has emerged as one of the most common post-translational modifications and as an important regulator of cellular signal transduction [139]. Methylation of non-histone proteins alters protein charges; hence, it may influence protein structure, activity, subcellular localisation, or interactions with binding partners [139, 141]. Histone proteins are methylated by histone methyltransferases (HMTs). In contrast, in non-histone proteins, two classes of enzymes carry out the addition of a methyl group to arginine's and lysine's: the protein arginine methyltransferases (PRMTs) and the protein lysine methyltransferases (PKMTs) [141-145]. Lysine methylation on non-histone proteins is usually associated with transcription and chromatin regulation [146] and occurs in K/R and/or S/T-K (serine/threonine-lysine) motifs on specific substrates, such as HSP90 (Heat Shock Protein 90) [147] and HSP70 (Heat Shock Protein 70) [148], both important to the stress response dynamics. However, herein, we will focus on arginine Methylation, as it is reported to regulate SAFB proteins [122].

Arginine is a positively charged amino acid, and its guanidinium group possesses the ability to form five hydrogen bonds (HB) with multiple molecules in water and within cellular membranes [149]. Therefore, when a methyl group is added to the arginine, its

positive charge is partially reduced. Hence, one of its HB sites is removed, consequently reducing or favouring its interactions with other molecules. There are nine different PRMT (PRMT 1 to 9) enzymes, which transfer a methyl group from S-adenosylmethionine (Ado Met) to guanidino nitrogen of arginine, resulting in S-adenosylhomocysteine (AdoHcy) and methylarginine. The Enzymes can methylate the arginine residues in three main types: type I, II or III, resulting in monomethyl arginine (MMA); asymmetric dimethylarginines (ADMA); and symmetric dimethylarginines (SDMA) (Figure 1. 3 Schematic image of the different types of arginine methylation (b). Adapted from Biggar, 2015

). All three types of enzymes methylate the terminal (or ω) guanidino nitrogen of arginine residues. Type I and type II enzymes catalyse the formation of an MMA intermediate, then type I PRMTs (PRMT 1, 2, 3, 4, 6, and 8) further catalyse the production of ADMA, while type II PRMTs (PRMT 5 and 7) catalyse the formation of SDMA (Figure 1. 4 Arginine methylation in mammalian cells. Type 1 and 2 PRMTs catalysed the transfer of the first methyl group from the Ado Met to the ω-Nitrogen atom of arginine, resulting in MMA. The subsequent reaction catalysed by type 1 PRMTs (1-4, 6 and 8) generates ADMA, whereas type 2 PRMTs (5 and 7) generates SDMA. Adapted from Zakrzewicz, D., 2012 .

) [142, 150, 151]. Whether demethylation occurred was a hotly debated issue until recently [152, 153]. If demethylation did not occur, the only way to undo this mark would be to degrade the protein and synthesise a new one. However, the discovery of the peptidyl arginine deaminase 4 (PAD4) enzyme in 2004 and the Jumonji domain-containing protein 6 (JMJD6) in 2007 [154, 155] presented proof of the reversibility of methylation. PAD4 can catalyse the demethylation reaction and convert arginine residues into citrulline, whereas JMJD6, which is a JmjC - containing iron and 2-oxoglutarate-dependent dioxygenase, de-methylate histone H3 at arginine 2 (H3R2) and histone H4 at arginine 3 (H4R3). The discovery of de-methylases proves the dynamic character of methylation, opening the possibility of a role for it in transient processes, such as the control of the stress response.

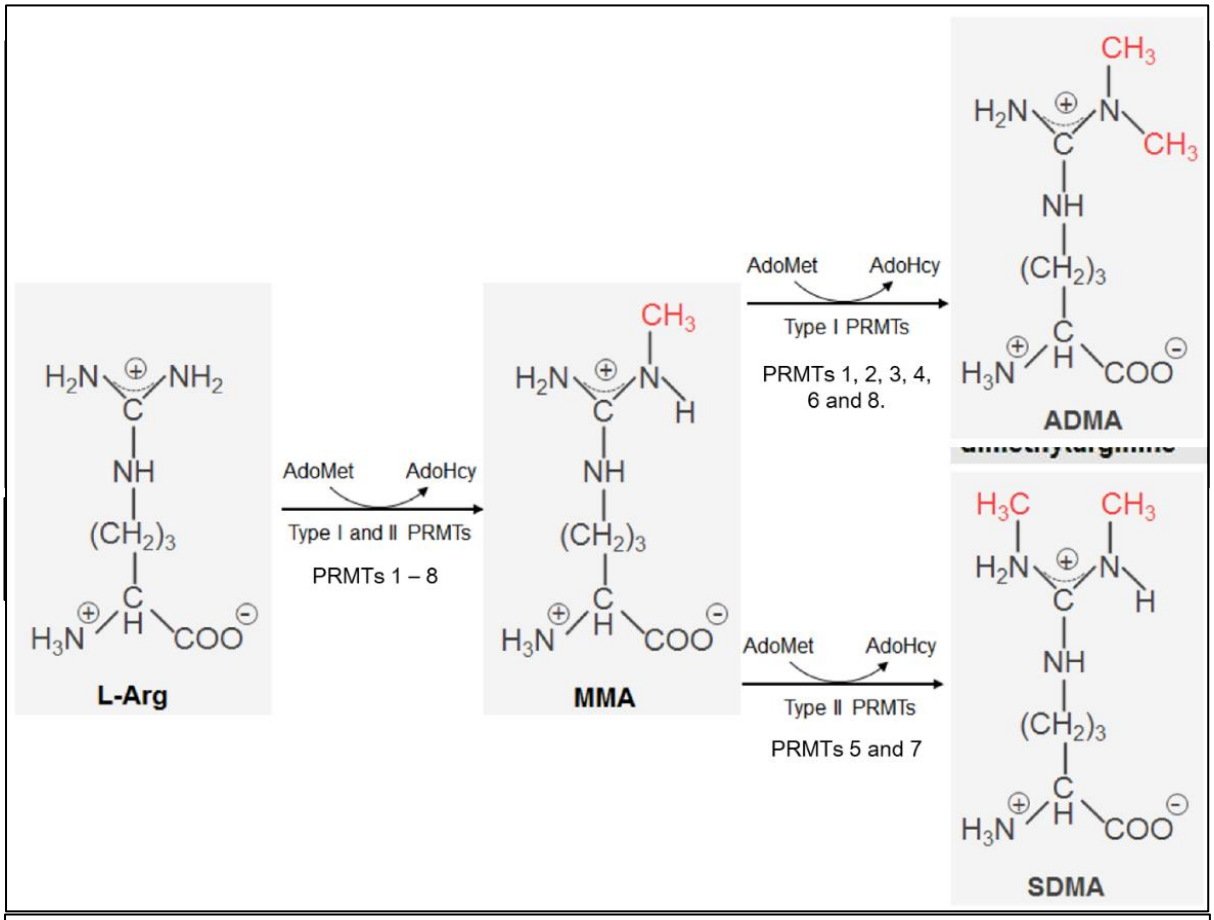

Figure 1. 4 Arginine methylation in mammalian cells. Type 1 and 2 PRMTs catalysed the transfer of the first methyl group from the Ado Met to the ω-Nitrogen atom of arginine, resulting in MMA. The subsequent reaction catalysed by type 1 PRMTs (1-4, 6 and 8) generates ADMA, whereas type 2 PRMTs (5 and 7) generates SDMA. Adapted from Zakrzewicz, D., 2012 .

PRMT enzymes prefer RG/RGG substrates [156, 157]. This RG/RGG motif is also the second most common RNA binding motif and it can be found in yeast and humans and is present in many proteins encoded by the human genome [158]. The RG/RGG motif is positively charged due to the the arginine residues, making it possible to interact with negatively charged compounds, such as DNA and RNA. In fact, the first evidence showing an interaction between an RG/RGG motif and nucleic acids was in 1992, with Kiledjian and Dreyfuss, where the authors showed repeated sequences of the RGG motif within SAF-A protein were required for SAF-A interaction with homopolymeric RNA [159]. Subsequently, several RG/RGG containing proteins were found to interact with RNA, for

instance, FUS, TAF15, hnRNP A1, SAF-A and SAM68 [160], which also interacts with SAFB1 [1, 14, 47, 159, 161]. SAFB family members also display RG/RGG motifs on its C-terminal end, mostly within the G-rich domain. Computational screening analysis for RG/RGG motifs in the human genome has found SAFB2 amongst the 88 human protein isoforms with a Di-RGG motif (RGG(X0-4)RGG), where X0-4 could be any amino acid sequence of four residues [160]. The same study identified SAFB1 isoforms one to four as containing a mono $RGG_{(X0-4)}$ RGG motif, among more than 1700 proteins found in the human genome, suggesting the participation of the G-rich domain on SAFB-RNA interactions. Additionally, the RG/RGG motif mediates protein-protein interactions as well. The Scd6 (suppressor of clathrin deficiency) in yeast interacts with eIF4G (Eukaryotic translation initiation factor 4G) through its RGG domain [162]. The arginine residues within the RGG motif of hnRNP A1 interacts with Rev, a protein adaptor for the nuclear export of HIV RNAs [163]. Finally, the RGG rich domain present in the protein nucleolin, a nucleolar phosphoprotein with a key role in DNA and RNA metabolism, can bide NS5B, a cleavage product from viral non-structural protein expressed during infection, via its RGG-rich domain [164]. Therefore, the evidence shows that RG/RGG motif is also implicated in protein-protein interactions, with possibly crucial roles in viral infection.

The function of arginine methylation on non-histone proteins are still under investigation. For instance, Boisvert and collaborators in 2002 reported that splicing reactions could be inhibited by SDMA antibodies in vitro [165], showing a role for arginine methylation in splicing for the first time. In 2004, Smith and collaborators showed that methylation inhibitors alter the RNA Helicase A (RHA) distribution. The enzyme typically undergoes nuclear translocation via its nuclear transport domain, and this function is impaired in the presence of methylation inhibitors [166], showing a role for arginine methylation in nuclear/cytoplasmic shutting. Moreover, the Mre11, a component of the MRN (Mre11/Rad50/Nbs1) complex that binds DSBs and tethers both DNA ends via non-homologous ends, and homologous recombination have an RGG motif methylated by PRMT1 [167-169]. When a global methylation inhibitor is added to mammalian cells, and Etoposide induces DNA damage, Mer11 is not recruited to the DNA strand break sites, and less γH2AX foci are formed [170]. Mer11 is involved in DNA double-strand repair, homologous recombination and telomere length maintenance [171], revealing a key role for arginine methylation in DNA repair and senescence

All SAFB family members are predicted to have arginine methylation sites (Figure 1. 5 Image scheme of SAFB1 and SAFB2 SUMOylation and Methylation sites according to the UniProt Database website.), with some of them already empirically identified [122]. According to studies using mass spectrometry to identify PTMs in human proteins, SAFB1 is methylated at three RG motifs (aa811 MMA, aa868, aa874 ADMA, and aa884 ADMA) and one RGG motif (aa868 ADMA), while seven RG/RGG motifs are predicted to occur in its canonical sequence. Computational analysis of SAFB1 isoforms sequences has shown that two out of its six RG motifs are located within the E/R-rich domain of all isoforms. In contrast, another four RG/RGG motifs are located on the G-rich domain [122]. Interestingly, an additional RGG motif is present in SAFB1 Isoforms four and five. Meanwhile, SAFB2 has two RGG motifs (aa897 and aa903, both MMA) already identified as di-methylated constitutively [122], although the protein sequence shows a total of five RG and tree RGG motifs, as shown in the image below (Figure 1. 5 Image scheme of SAFB1 and SAFB2 SUMOylation and Methylation sites according to the UniProt Database website.). Moreover, SLTM has seven RG sites, one of them already validated within the C-terminal end of the protein. Nevertheless, it is important to remember that the E/R-rich domain is responsible for many important protein-protein interactions [1, 9].

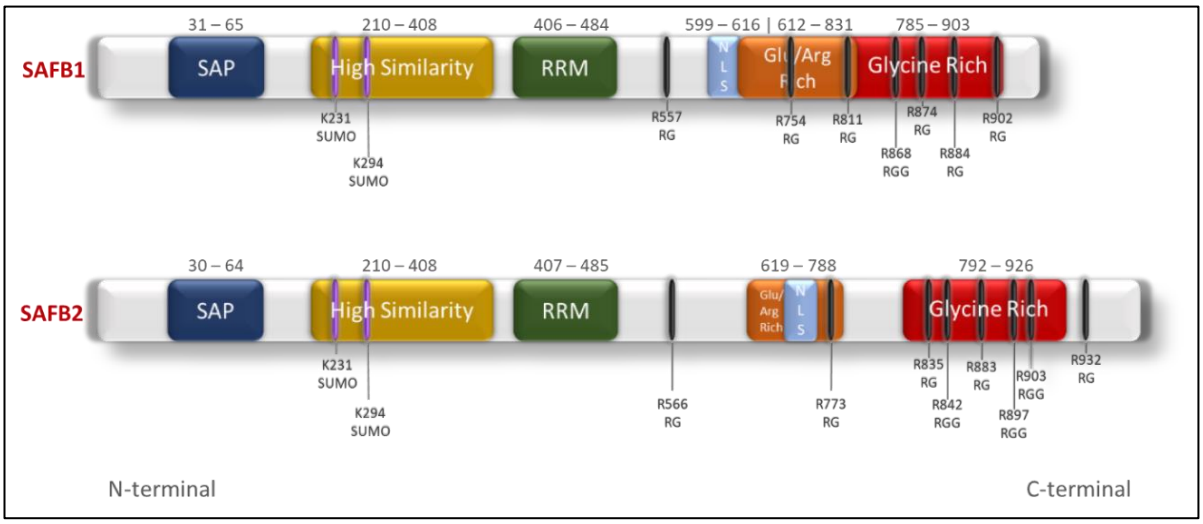

Figure 1. 5 Image scheme of SAFB1 and SAFB2 SUMOylation and Methylation sites according to the UniProt Database website.

In summary, SAFB proteins have functions in multiple cellular processes, namely RNA processing, chromatin organisation and stress response.

1.6 Stress Response

The Heat Stress response:

Eukaryotic cells have adapted and evolved to survive various environmental insults, such as radiation, toxic molecules, heavy metals, osmotic stress and a broad range of elevated temperatures. Higher temperatures can cause the denaturation and unfolding of proteins, entanglement and non-specific aggregation, leading to loss of cell cycle control, cell damage and/or death [172]. Therefore, cells developed a means of protecting themselves from various environmental insults, which was termed the Heat Shock/Stress Response (SR) [173].

During the SR, constitutive transcription is inhibited to prevent the synthesis of abnormal proteins while the cell increases the synthesis of protective Heat Shock Proteins (HSPs) [174]. The HSPs are chaperones that bind to newly synthesised, misfolded or damaged proteins, important for maintaining the protein homeostasis and preventing entangling and aggregation of proteins during stress [reviewed in [175, 176]]. Heat Shock Factor 1 (HSF1) is the main player in the SR, being implicated in the up-regulation of HSPs upon heat stress [177] and driving the transcription of pericentromeric satellite DNA (SATIII) at the 9q12 chromosome [178, 179]. The SATIII long-non-coding RNA transcripts are not translated into protein, and they remain in the vicinity of their transcription sites, acting as a sponge to transcriptions factors, competitively inhibiting transcription [180]. Moreover, another important outcome of the stress response is the upregulation of the HSP genes [181]. When activated, HSF1 binds to the promoters of HSP genes, specifically to nGAAn pentamers, and recruits co-factors to increase their expression [182]. Thus, the cell undergoes a dramatic shift in its metabolism to cope with environmental stress and overcome the adversities. Moreover, this process needs to be finely regulated to prevent cell death or loss of cell cycle control.

1.6.1 Nuclear stress bodies and HSF1

Heat Shock Factor 1 is one of four members of the HSF family (HSF 1-4) in vertebrates [177]. The genome sequence encoding the protein is located at the 8q24.3, in which transcription activation produces only one isoform with a 14-exon count (Gene ID: 3297). The newly formed protein has 529 amino acids and a molecular weight of 57.1 kDa before PTMs (ACCESSION NP_005517). Under normal conditions, HSF1 is kept in a non-DNA-binding confirmation as an inactive monomer complexed with HSP90 [183, 184] mostly within the nucleus excluded from the nucleoli [185, 186]. HSF1 is released from the HSP90 complex during the stress response and forms either homotrimers or heterotrimers with HSF2 [187]. The trimers are then activated via hyper-phosphorylation on Serines 121, 230, 292, 303, 307, 314, 319, 344, 363, 419, and 444, in which Ser326 seems to play a more important role. This activation is reversible, and HSF1 returns to an inactive state during post-heat stress recovery [188].

The Active HSF1 trimers rapidly bind and aggregate specifically at the 9q12 pericentromeric region of chromosome nine and secondary sites on chromosomes 12 and 15 [179]. The immunofluorescence against HSF1 before and after heat stress is seen in Figure 1. 6 Nucleus of Human fibroblasts stained using indirect immunofluorescence for HSF1 following a heat stress. (A) Cell under normoxic conditions; (B - D) Cells heat-shocked at 42°C showing the three types of HSF1 nSBs: (B) cluster of small granules; (C) large solid foci and (D) two unique punctate signals, described by Jolly et al. Bars: 5 μm.

, and Figure 1. 7 shows the HSF1 accumulation pattern seen in mammalian cells. These sites are called nuclear Stress Bodies (nSBs), and they are different from other previously described sub-nuclear structures such as nuclear speckles, Cajal bodies and PML (promyelocytic leukaemia) bodies [189].

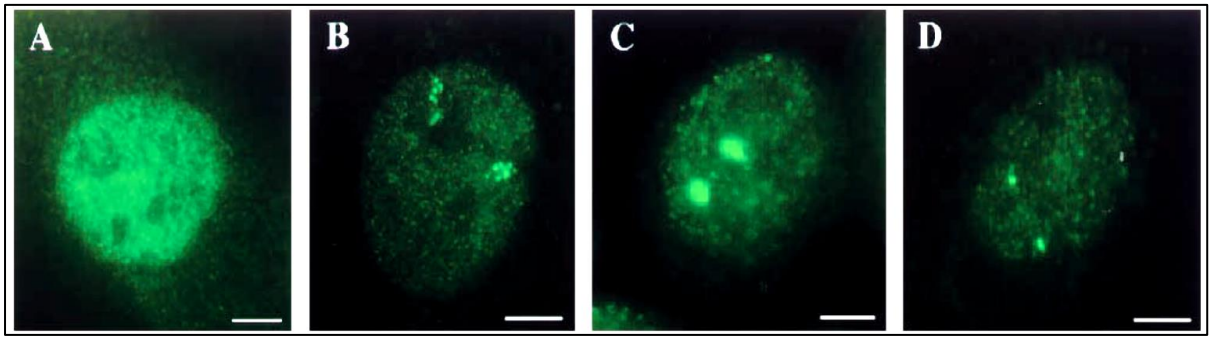

Figure 1. 6 Nucleus of Human fibroblasts stained using indirect immunofluorescence for HSF1 following a heat stress. (A) Cell under normoxic conditions; (B - D) Cells heat-shocked at 42°C showing the three types of HSF1 nSBs: (B) cluster of small granules; (C) large solid foci and (D) two unique punctate signals, described by Jolly et al. Bars: 5 µm.

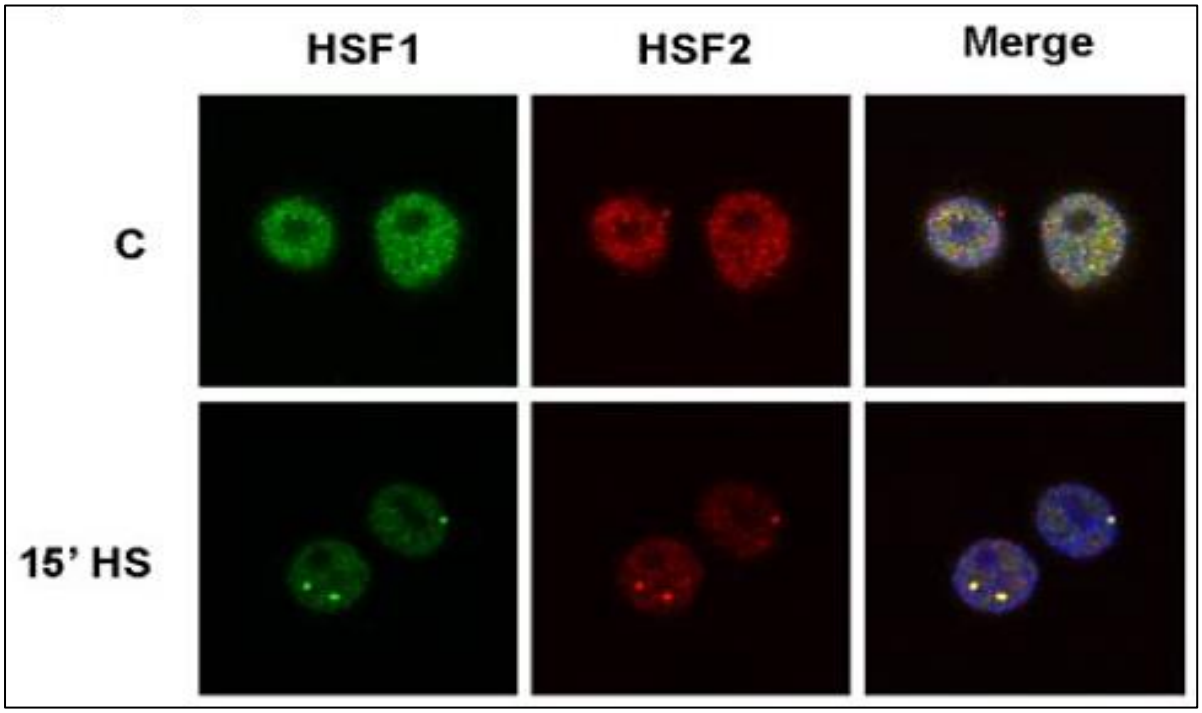

Figure 1. 7 HeLa cells stained for immunofluorescence against HSF1 and HSF2 15 minutes after induction of heat shock at 42°. Image taken from Sandqvist, 2009.

For a long time, it was believed that these accumulation sites of HSF1 observed by immunofluorescence were located to the promoters of Heat Shock Elements (HSEs), driving the activation of HSPs, such as HSP70 and HSP90 [181, 182]. However, in 2002, Jolly and collaborators published a study proving that the genomic binding/accumulation sites for HSF1 in primary human cells (fibroblasts) were regions of repetitive DNA within the centromeric portion of chromosome nine (9q11 – q12), which correspond to Satellite III DNA sequences [178]. The study showed that in primary human metaphase, fibroblast-spread HSF1 always formed two distinct foci localised in the region below the centromere of Chromosome 9, one in each allele. Moreover, the only requirement for foci assembly was the trimerisation of the protein and direct biding DNA [178].

HSF1 recruitment to SATIII loci mediates the assembly of nSBs following stress. nSBs become visible only five minutes after a 5°C increment in temperature in normal skin fibroblasts. The percentage of cells showing these structures increase with time, displaying a peak of 90% at 60 minutes of HS. After 1 hr of recovery at 37°C following heat stress, most of the cells are positive, although this figure decreases by 2.5% after 6 hr of recovery. It also has been reported that the number of nSBs formed upon HS stress decreases with the increase of cell passages. Therefore, there might be a relationship between senescence and HS response [41].

Further studies from Jolly`s group also showed that in tumour cells (such as HeLa and HCT116), HSF1 binds to additional secondary loci in 14 different chromosomes, always in regions encoding repetitive Sat II and III sequences[41]. Tumour cells express abnormal amounts of several contents, including HSF1. Hence, the ability of HSF1 to bind secondary sites was co-related to the amount of protein available for the cells. Secondary binding sites of HSF1 were found in chromosomes 1, 5, 7, 9, 10, 13, 14, 15, 16, 17, 21, 22 and Y, and they were shown to be transcriptionally active. Authors, therefore, defined the 9q11-q12 recognition region by HSF1 in both primary and tumour cells as a primary binding site of HSF1. The subsequent HSF1 binding only found in tumour cells was termed secondary biding sites.

Once directed to either primary or secondary biding sites, HSF1 recruits RNA polymerase II, which drives the transcription of the satellite non-coding RNA. Despite the recruitment

of RNA pol II to the HSF1 biding sites, none of the adjacent genes had their expression increased unless the gene were somehow related to the heat stress response [41].

1.6.2 lncSATIII

Satellite III DNA occurs at pericentromeric regions and contains a repeated GGAAT sequence located at the 9q12 chromosome region. Transcription of SAT III long non-coding RNA is increased upon environmental stress such as heat stress, heavy metals, osmotic pressure, ibuprofen, MG132, puromycin and others [78, 189, 190]. Activated trimers of HSF1 binds to the SATII sequences driving the recruitment of RNA polymerase II and subsequently the transcription of the repeats into a long non-coding RNA (lncSATIII). These transcripts remain in their site of synthesis, accumulating and recruiting splicing and transcription factors, such as SRSFs (serine/arginine-rich proteins), hnRNPs, CREB binding protein [1, 41, 77, 179, 190-192]. HSF1 and the lncSATIII co-localise within the nSBs, and the absence of HSF1 inhibits the synthesis of the transcripts. Interestingly, depletion of the lncSATIII results in elimination of nSBs and promotes splicing of 533 retained introns during thermal stress recovery [20] and partially relieves the transcriptional repression seen in the stress response, whereas enforced expression of it induces the formation of nSBs and transcriptional repression even without stress induction [190].

1.6.3 HSPs

Important proteins involved in the HSR are HSP70 and HSP90. These protein-chaperones are involved in folding newly synthesised proteins and the refolding of misfolded or aggregated proteins [193]. They are both upregulated in stress to prevent protein damage. As newly synthesised proteins emerge from the ribosomes, Hsp70 recognises sequences of hydrophobic amino acid residues and interacts with them. This not only prevents incomplete proteins from aggregating and being rendered non-functional but also avoid the release of damaged proteins into the cytoplasm upon stress [193-195].

Aims of this study

SAFB1 has been extensively studied in comparison to SAFB2 and roles in transcription, cell homeostasis and the regulation of the cellular stress response have been found. The

high degree of homology between these paralogues has led to the assumption that SAFB2 has similar functions to the more highly investigated SAFB1. This thesis describes experiments using reagents and molecular techniques to examine the specific roles of SAFB1 and SAFB2. Specifically, for the first time we aimed to investigate the role SAFB1 and SAFB2 plays in nuclear stress body formation. In addition, I aimed to investigate whether SAFB1 and SAFB2 proteins underwent differential arginine methylation following stress and aimed to identify novel SAFB1 and SAFB2 protein interaction partners.

In the following results chapters I describe experiments that show SAFB1, SAFB2 and SLTM have distinct recruitment patterns following thermal stress. SAFB2 was found to co-localise with 87% of canonical HSF1 nSBs, a major maker of heat stress, as opposed to 6.7% of co-localisation between SAFB1 and canonical HSF1 nuclear bodies. SAFB2 also shows higher co-localisation HS induced lncSATIII transcripts then SAFB1. Moreover, SAFB2 depletion increased the expression of lncSATIII transcripts following heat stress, an effect that was not seen with either SAFB1 knockdown or double SAFB1/SAFB2 knockdown. I also investigated whether symmetric arginine di-methylation and SUMOylation of SAFB1/2 proteins was regulated by stress. The global inhibition of methylation decreased the formation of canonical and secondary HSF1 nSBs following HS, increased large SAFB1 aggregate formation, and had no effect on SAFB2 distribution. Whereas prevention of symmetric arginine di-methylation significantly increased the formation of canonical HSF1 nSBs but produced no effect on the distribution of any SAFB family member. Moreover, SAFB1 and SAFB2 mutants containing mutated SUMOylation and arginine methylation sites showed a delay in their recruitment to HSF1 defined nSBs during HS. Finally, proteomic analysis of proteins co-immunoprecipitated with SAFB1 and SAFB2 was used to identify novel interaction partners for SAFB1 and SAFB2 proteins. SAFB1 and SAFB2 both co-immunoprecipitated Ago2, a novel interacting partner not previously reported in the literature, which has recently been linked to Huntington's disease [100], corroborating our recent findings of abnormal expression of SAFB1 protein in striatal neurons in *postmortem* brain tissue of Huntington's patients [95]. Furthermore, Ingenuity Pathway Analysis (IPA) was performed to explore novel cellular processes and SAFB1 and SAFB2 proteins were shown to be highly involved in miRNA processing, eIF2/ISR signalling, the unfolded protein response and Huntington's Disease signalling. Together, these data reinforce a key role of SAFB1

and SAFB2 in regulating critical cellular processes such as miRNA processing, cellular adaptation to stress and neurological disease development.

Chapter 02: Methods

2.1 Overview

Here we offer a description of most of the material and methods used in this thesis. Certain approaches are restricted to specific chapters; therefore, they will be mentioned here and described in detail later in this thesis.

In Chapter 3, the aim was to detect the subcellular localisation of SAFB family members (SAFB1, SAFB2 and SLTM) and their recruitment upon stress conditions on in-vitro neuronal models. We used combined techniques such as ICC, high content imaging and modular Image analysis to access SAFB proteins recruitment.

In Chapter 4, overexpression of enhanced green fluorescent protein (EGFP)-tagged SAFB1 mutants deletion and EGFP-SAFB Methylation mutants construct were used to identify the effects of these deletions and modifications/changes on SAFB recruitment under normal and stress conditions. The SAFB1 Deletion mutants were generated and kindly donated by Dr Mathias Altmeyer, as described in [11]. Additionally, protein methylation inhibitors Adenosine dialdehyde (AdOx) and GSK591 were used to evaluate the effects of protein Methylation impairment during the heat shock response.

The main goal in chapter 5 was to evaluate the recruitment of SAFB1 and SAFB2 in primary human cells and validate potential biding partners for SAFB1 and SAB2, such as Argonaute2.

2.2 Cell culture

All cells were grown in T75 flasks at 37°C in 5% CO2 atmosphere. For the experiment's purposes, cells were plated and grown at appropriate densities.

2.2.1 Human neuroblastoma cell line (SH-SY5Y cells)

The SH-SY5Y (ATCC® CRL-2266™) is a human cell line from Bone Marrow that developed into a Neuroblastoma. SH-SY5Y cells exhibit a close to normal karyotype (modal number = 47), with a unique marker trisomy of 1q [196]. The cells exhibit an epithelial morphology when cultured in adherence, although they could also be cultured in suspension. Interestingly, SH-SY5Y cells are reported to exhibit moderate levels of

dopamine beta-hydroxylase activity [197], bringing them close to a neurological cell type. Therefore, this cell line was chosen to perform most of the experiments conducted and described in this thesis.

SH-SY5Y cells were maintained in Dulbecco's modification of Eagle's medium / F-12 (1:1) (DMEM-F12) nutrient mixture, supplemented with 10% (v/v) foetal bovine serum (FBS) and L-glutamine (2mM), all purchased from Sigma-Aldrich, UK. Antibiotics were avoided due to their interference with transfection procedures and to avoid hidden contaminations. Cells were passaged once a week in a ratio 1:10-1:12 and used for experiments between passages 14 and 22. Passaging and/or plating of the cells were carried out by washing cells with room temperature (RT) PBS, followed by 10% (v/v) trypsin- EDTA solution digestion harvested in a pre-warmed 3x (v/v) medium in order to stop the trypsin activity. The cell suspension was centrifuged at 1,2000 x g for two minutes at room temperature, and the cell pellet was resuspended in 8-10 ml of pre-warmed medium. The number of cells was counted before being plated in flasks or culture dishes at various densities and kept at 37°C in a 5% CO2 atmosphere before experiments.

2.2.2 Human cervical cancer cell line (HeLa)

HeLa cells were maintained in Dulbecco's Modified Eagle Medium (DMEM) supplemented L-glutamine (2mM) and 10% FBS (all Sigma-Aldrich). Cells were passaged twice a week as described above (2.2.2) with complete media changes.

2.2.3 Human Fibroblasts

Human skin fibroblasts were obtained from a 49-year-old male with no history of clinical diagnosis or neurodegeneration by Dr Lucy Crompton, who has kindly yielded the cells for this thesis. The fibroblasts were defrosted and kept in DMEM media supplemented with Glutamax (1:100, Life Technologies), 10% FBS, Non-essential Amino acids (1:100, Life Technologies) and Penicillin/Streptomycin (1:100, Life Technologies). Before seeding/passaging, flasks, plates or dishes were coated with 0.1% Gelatine (Sigma) in PBS.

2.3 Glass coverslips treatment poly-D-lysine coating

Specific immuno-fluorescent protocols required the cells to be plated on coverslips to be mounted into slides. For this, circular borosilicate covered glasses, between 0.13 - 0.16 mm of thickness (VWR), were kept in 70% Ethanol 24hrs before their use. Prior to cell plating, coverslips were dipped into 100% ethanol, flamed-sterilised and laid in 24 or 4 well plates (Cellstar Greiner Bio-one, UK). 0.5 mL of poly-D-lysine (0.1mg/mL, Sigma-Aldrich, UK) was added to each well of the culture plates and stored either at 4°C overnight or 37°C for 1-2 hrs. Coverslips were then washed in sterilised, deionised water to ensure complete removal of poly-D-lysine as it is toxic to the cells. Wells were left to dry before cells were plated in appropriate densities.

2.4 Cellular stress by Heat Shock (HS):

In order to induce heat stress the cell environment must be raised to a temperature higher than homeostasis (>37°C) but not not exxed 43°C which can lead to long term irreversible cell damage in different mammal tissues [198, 199]. Our for the last 1o years has used a standardised method to induce in vitro heat shock, using a temperature high enough to induce changes but not enough to produce deleterious effects. Hence, increasing the cell media temperature to 42°C, for 1hr. This allows results to be compared with previous work in our laboratory and is comparable to that used by groups around the world [1, 104, 191, 200-202]. To ensure the accuracy of the temperature, digital thermometers were placed in wells within the incubator, and many members of our laboratory found this treatment induced the expression of heat shock proteins and the maximum formation of nSBs.

This, an empty heat block was placed inside the incubator in 5% CO2 atmosphere and set to 43.8° Celsius. Once the desired temperature was achieved, the target cells (previously plated and grown in appropriated densities on plates/dishes and submitted or not to chemical treatments) were laid on the empty heat block, and media temperature was monitored to confirm achievement of 41.8 to 42 °C degrees, which in average took 20 minutes. Cells were then left on the heat block for another 60 minutes for HS induction with constant temperature monitoring. For most of the experiments, three time-points were applied: (a) No Heat Shock (NHS), in which cells were kept under physiological

conditions (37°C) at all times; (b) Heat Shock No Recovery (HSNR), in which cells were submitted to 60 minutes of heat stress and fixed/harvested immediately after this period; or (c) Heat Shock + one hour of Recovery (HS1hr), in which cells were submitted to 60 minutes of heat stress and then placed in the incubator out of the heat block and left to recover for 60 minutes at 37°C before fixing/harvesting. In cases where different HS times were used, the conditions will be specified.

2.5 Plasmid Overexpression

Plasmid overexpression was carried out in different sections of this document, using specific plasmid vectors depending on the experiment purposes. Each plasmid used will be specified in its results sections. However, the protocol used remained the same across the experiments.

Cells were plated and grown for 24 hours as described previously, in suitable plates and densities. When appropriate, a complexed solution was made, comprised of Opti-MEM (Gibco), 500 ng of plasmid DNA and 1.5 μL Lipofectamine 2000 /well (Thermo Fisher Scientific, UK) as a transfection reagent. The solution was prepared following the Lipofectamine 2000 manufacturer instructions and added to the wells by gently dripping it across all the cells surfaces. Cells were kept in the incubator at 37°C in 5% CO2 atmosphere for 48 hours prior to any other experiments.

2.6 Live Imaging: IncuCyte Zoon

In order to determine the time course of SAFB nSB formation, EGFP-tagged plasmids encoding for either SAFB1, SAFB2 or SLTM proteins under a CMV promoter were overexpressed in HeLa Cells, and images were acquired using the IncuCyte Zoom System. The IncuCyte Zoom is a microscope system inside a CO2 incubator at 37°C that provides live images of cells under physiological conditions. The system holds an automated time-lapse image capture in multiple positions with 4x (0.2 NA), 10x (0.3 NA) or 20x (0.45 NA) lenses and offers phase contrast and Green/Red fluorescence imaging. Hela cells were cultured as described before and plated in suitable densities in 24 well-plates 24 hours prior to the transfection. Transfection was carried out as described previously. Before heat stress induction, plates containing transfected cells were placed in the IncuCyte Zoom and images were acquired as a non-heat shock control. The plates

were then removed from the system, heat shocked as described previously and placed back in the same positions to allow cell track. From now on, images were acquired every 30 minutes, under phase contrast and green fluorescent filter for visualisation of the EGFP tag, up to 6 hrs. This experiment was not repeated due to its low-resolution proprieties.

2.7 Small interference RNA (siRNA) and Knockdowns:

Knockdown of SAFB proteins is a recurrent technique performed in this thesis. Therefore, this section will provide a general guideline of how protein knockdown was achieved. Specific information will be provided in each result section in further chapters.

Cells were plated and grown in appropriate densities for 24 hours prior to transfection with siRNAs. Table 3.2 shows the OnTarget Plus pools purchased from Dharmacon/Horizon, UK (Figure 2. 1). For each knockdown, 60 pM of siRNA against either SAFB1, SAFB2, SLTM, or a non-targeting control siRNA (NTC), were used in each well. In addition, a fourth group was added, where cells were transfected with 30 pM of SAFB1 siRNA plus 30 pM of SAFB2 siRNA to produce a double knockdown.

Lipofectamine RNAiMax (Thermo Fisher Scientific, UK) was used as a Transfection Reagent coupled with Opti-MEM Reduced Serum Media (ThermoFisher, UK), following the manufacturer instructions. The complexed siRNA-Lipofectamine-Opti-MEM was added to each well by gently dripping it across all the cells surfaces, and cells were incubated at 37°C/ 5% CO2 for 48 hours prior to protein/RNA harvesting.

Target	siRNA pool	Concentration on cells [pM]	#Catalogue	Manufacturer
SAFB1	onTARGET plus Human SAFB-SMARTpool	60	L-005150-00-0010	Dharmacon/Horizon
SAFB2	onTARGET plus Human SAFB2- SMARTpool	60	L-020373-01-0010	Dharmacon/Horizon
Double KD	SAFB1 pool + SAFB2 pool	30 + 30	L-005150-00-0010 L-020373-01-0010	Dharmacon/Horizon
SLTM	ON-TARGETplus Human SLTM siRNA	60	L-020373-01-0010	Dharmacon/Horizon
NTC	onTARGET plus Non-targeting Pool	60	D-001810-10	Dharmacon/Horizon

Figure 2. 1 List of siRNAs against SAFB1, SAFB2, SLTM and non-targeting control siRNA purchased from Horizon, UK (former Dharmacon)

2.8 Protein analysis

2.8.1 Cell lysis and protein collection

A minimal volume of cold RIPA buffer (see 2.9.4 Solutions below) supplemented with protease inhibitor (Complete Mini; Roche) was added to each well to lysate cells and stabilise the protein. The lysate was transferred to a new fresh tube and triturated using a syringe and needle before being centrifuged at 10,000 x g for 20 minutes at 4ºC. The supernatant was collected into fresh tubes and forwarded either to protein analysis or storage at -80ºC. To measure the protein content of each sample, albumin standard dilutions were prepared using the bicinchoninic acid (BCA) protein assay kit (Thermo Fisher Scientific, UK). Standard dilutions and samples were distributed into separate wells in a 96-well plate that contained BCA reagents. Protein concentrations were measured by quantifying the absorbance at 550 nm using a plate reader (Glomax Multidetection system, Promega) regarding the standard dilutions.

2.8.2 Electrophoresis

8% acrylamide SDS gels were prepared (see 2.9.4 Solutions below) and set up in a Mini Protean II apparatus (Bio-Rad, Hercules, USA). Butanol was added upon the non-polymerised gel to aid in settling, level and prevent air bubbles and removed shortly after polymerisation. The stacking gel solution was then poured on top of the resolving gel, and a comb was inserted between the glass plates to create wells for sample loading. Gel was left at room temperature to be polymerised. Once the gel was ready, it was assembled in a Mini Trans-Blot® tank (Bio-Rad) filled with running buffer (see recipe below). The protein lysates were mixed with working buffer (2 x SDS reducing sample buffer with Dithiothreitol [DTT]) at desired protein masses and prepared by boiling them at 100 ºC for 5 minutes followed by 5 minutes incubation on ice prior to gel loading, along with 1kb protein standard (NEB, UK). Samples were run at 100V until the desired separation of the standards was achieved.

2.8.3 Transfer/staining and detection of proteins

Protein was transferred to a Polyvinylidene difluoride (PDVF) membrane using the Trans-Blot Turbo Transfer System (Bio-Rad) and Trans-Blot Turbo Mini PVDF Transfer Packs

(Bio-Rad). The transfer was carried out at 1.3A for 10 minutes. Once the protein was successfully transferred onto a membrane, it was then blocked in PBS-Tween containing 5% milk (Dried skimmed milk, Marvel) for 1 hour. Primary antibody incubation of the membrane was carried out overnight at 4°C with gentle rocking. After washes, the membrane was incubated with horseradish peroxidase (HRP)-conjugated secondary antibodies at room temperature for 1 hour. The membrane was washed, and bands were detected using enhanced chemiluminescence (ECL) reagent (Pierce) by exposing it to an autoradiographic film (Amersham, GE Healthcare. Germany). The film was developed using an auto processor (Kodak). The relative protein expression was calculated by densitometry using ImageJ and normalised to Alpha tubulin to control for loading.

2.8.4 Solutions

- RIPA: 50mM Tris-HCl, pH7.4; 150mM NaCl; 1% Triton X-100 (v/v); 1% sodium deoxycholate (w/v) and 0.1% sodium dodecyl sulphate (SDS) (w/v).
- 8% SDS Gel: 8% Acrylamide, 375 mM Tris pH 8.8, 0.1% Sodium Dodecyl Sulphate (SDS), 0.1% Ammonium Persulfate (APS), 0.06% N,N,N′,N′-Tetramethyl ethylenediamine (TEMED).
- Staking Gel (6%): 6% Acrylamide, 125 mM Tris pH 6.8, 0.1% Sodium Dodecyl Sulphate (SDS), 0.1% Ammonium Persulfate (APS), 0.04% N,N,N′,N′-Tetramethyl ethylenediamine (TEMED).
- 2 x SDS reducing sample buffer: 50nM Tris-HCl, pH 6.8; 100 mM DTT; 2% SDS [w/v]; 0.1% bromophenol blue [w/v] and 10% glycerol [v/v].
- Running buffer: 25 mM Tris, 190 mM Glycine, 0.1% SDS.
- 2xSSC: 17.53 g NaCl, 8.82 g sodium citrate, 800 ml dH2O.

2.9 Immunocytochemistry

For conventional and/or confocal microscopy, cells were plated on PDL-coated coverslips, whereas for high content analyses (Opera system), they were plated on glass-bottomed 24 well-plates (CellVis, USA). Following the appropriate treatments, cells were washed in PBS and fixed in 4% paraformaldehyde (PFA) for 15 minutes at 37°C. Cells were blocked and permeabilised in 0.3% Triton X-100 with 5% donkey serum in PBS for

1 hour. Primary and secondary antibodies were added to each well in a 1:200 dilution in 1% Donkey serum PBS. Primary antibodies were incubated overnight at 4°C, followed by 3x10 minutes washes with PBS. Fluorescent conjugated secondary antibodies were incubated for 2 hours at room temperature and washed 3x10 minutes in PBS to remove excess antibodies. All incubation steps were carried out in gentle rocking. Finally, cells were mounted with either VECTASHIELD® Antifade Mounting Medium with DAPI (Vector Labs) when plated on Celli's glass-bottomed well-plates, or incubated for 10 minutes with Hoechst solution and mounted on glass slide by using Mowiol mounting medium (Calbiochem, Millipore, UK) with 2.5% w/v 1,4-diazabicyclo-[2.2.2] octane (DABCO; Sigma-Aldrich, UK). Figure 2. 2 shows a list of the commonly used antibodies across the experiments,

Antibody	Company	Catalogue Number
SAFB1 (rabbit)	Bethyl	A300-811A
SAFB2 (rabbit)	Bethyl	A301-112A
SLTM (rabbit)	Bethyl	A302-835A
HSF1 (rat)	ENZO Life sciences	ADI-SPA-950
SRSF1 (mouse)	Thermo Fisher Scientific	32-6400
donkey anti rabbit IgG Cyanine 3	Jackson Laboratories	711-165-152
donkey anti rat IgG Cyanine 3	Jackson Laboratories	712-165-150
donkey anti rabbit IgG Cyanine 2	Jackson Laboratories	711-225-152
donkey anti mouse IgG Alexa Fluor 488	Thermo Fisher Scientific	R37114

Figure 2. 2 List of the most common primary and secondary antibodies used in the immunocytochemistry.

2.10 RNA-Fluorescent *In Situ* Hybridization

The RNA-Fish technique was performed in PDL-coated coverslips of 24 Glass bottom wells plates (CellVis, USA). Cells were plated, grown, and treated accordingly and then fixed in 4% PFA as described before, permeabilised with PBS-0.5% Triton-X for 5 min on ice and washed in PBS and 2xSSC buffer. The washing solution was replaced by a Hybridisation buffer (2× SSC buffer, 25 % formamide, 5× Denhardt solution, 10 % dextran sulphate, 1 ug/mL yeast tRNA and 5 ng/mL probe) and incubated overnight at 42°C in a

humid chamber. Figure 2. 3 lists the biotinylated probes used for the SATIII transcripts hybridisation

Probe	Sequence	Label	Company
Direct SATIII	GGAATGGCATGGATTGGAAT	Biotinylated	Sigma
Indirect SATIII	ATTCCAATCCATGCCATTCC	Biotinylated	Sigma
pBR3222	CCGGGAAGCTAGAGTAAGTAG	Biotinylated	Sigma

Figure 2. 3 shows the biotinylated RNA probes used on RNA-FISH Technique

Following incubation, wells were rewashed in 2xSSC buffer and blocked for at least 30 minutes at 37°C with 1x ISH Blocking solution (Vector Labs, USA). The blocking solution was then replaced by Fluorescein Avidin D (5 µg/mL) diluted in blocking solution and incubated for another 30 minutes at room temperature. In order to increase the fluorescence signal, an additional 30-minute incubation step with Biotinylated Anti-Avidin D solution (5 µg/mL – Vector Labs, USA) was performed. Plates were then incubated again with Fluorescein Avidin D (5 µg/mL) for 30 minutes. Figure 2. 4 shows the list of antibodies used in the RNA-FISH technique. Solution changes were all intercalated with 2x 3-minute washes with blocking solution. The final step was comprised of 2x 5-minute washes with 4xSSC+0.1% Tween 20 solution. Cells were then forwarded to ICC as described before. The experiments were repeated 3x separately.

Antibody	Concentration	Company	Cat. Number
Fluorescein Avidin D	5 µg/mL	Vector Labs, USA	A-2001
Biotinylated Anti-Avidin D	5 µg/mL	Vector Labs, USA	BA-0300

Figure 2. 4 shows the list of antibodies used on RNA-FISH Technique

2.11 Imaging

Images were captured using: i) confocal laser scanning microscopy (CLSM) (Leica SPE; Leica Biosystems, Germany); ii) OperaLX high-throughput microscope system

(PerkinElmer, USA) with a fully automated 1.2 NA 60x water immersion objective and IR autofocus.

Confocal images were further processed using Fiji/ImageJ software. An average of 30 images was taken of each well, using the same parameters for different antibodies through a 60x water immersive lens. Due to the higher magnification, the probability of imaging empty spaces is considerable. Therefore, images were sorted before analysis and a minimal of 10 images per well were included in each result set.

2.12 Image analysis

To characterise the formation/assembling of SAFB stress bodies during the stress response, we used the Fiji/ImageJ workflow automation plugin, MIA (version 0.10.8) [203, 204]. Before processing, 3D image stacks were reduced to 2D using a maximum intensity z-axis projection. Nuclei were detected using the blue fluorescence channel. First, the noise was suppressed using a median filter (σ = 2px) followed by the heterogeneous microscope illumination profile correction using rolling ball background subtraction (σ = 50px). A second median filter further removed noise (σ = 2px). The resultant image was binarised using a modified 3D Phansalkar local threshold [205] with a 0.6x intensity multiplier applied to the threshold to reduce the calculated threshold value systematically. Spatially proximal nuclei that had become merged into a single object during binarisation were split using a distance-based watershed transform (a preceding hole filling step was used to prevent over-segmentation) [206]. Nuclei were identified in the binarised image as contiguous regions of foreground-labelled pixels [207]. These nuclei were subject to a size filter based on a broad range of expected nuclear cross-sectional areas (50 μm^2 < area < 500 μm^2).

Stress bodies were detected in the green and red fluorescence channels using similar processes. The stress body signal was enhanced using pixel-classification with the ImageJ/Fiji WEKA plugin [206]. The Trainable Weka Segmentation Tool (TWST) was used to train the ImageJ software to recognise nSBs in different channels. The TWST is a learning tool that creates an algorithm (^{alg}) based on examples that identify the objects according to the user`s input. Several images of equal staining across different time points and N numbers are loaded into the TWST, where the user manually selects the

background, the nSBs and the noise as shown in Figure 2. 5, training the software to identify the structures. The same process is repeated for each staining set. Therefore, five algorithms were created to identify nSBs, corresponding to HSF1, SAFB1, SAFB2, SLTM and lncSATIII staining. Combinations of TWST algorithms ($HSF1^{alg}$ + $SAFB1^{alg}$ / $HSF1^{alg}$ + $SAFB2^{alg}$ / $HSF1^{alg}$ + $SLTM^{alg}$) are then loaded into the MIA, which used these algorithms to analyse the entire set of images.

The resultant probability images were binarised at fixed probability thresholds of 0.375 for green stress bodies and 0.75 for red stress bodies. Holes in the binarised objects were removed prior to identifying individual objects using connected components labelling [207]. Stress body objects were subject to a size filter (0.1 μm2 < area < 10 μm2). Any stress bodies not coincident with a previously detected nucleus were excluded from further analysis. Overlapping stress bodies were identified and moved to a separate object collection. Finally, the green, red and overlapping stress body collections were split into two categories based on a size threshold of 0.7 μm2. Size parameters for nSBs were adapted from Joly and collaborators [185]. The analysis is then converted into an excel spreadsheet containing all the data extracted from the images. The numerical results were then be used in statistical analysis to formulate averages and plot graphs comprised of the important data from the images. One-Way Annova statistical test was applied, followed by Bonferroni`s multiple comparison tests to obtain statistical relevance. An average of 30 images was acquired per well, in triplicates unless stated otherwise, in three separate experiments. At least ten images/well were loaded into the analyses.

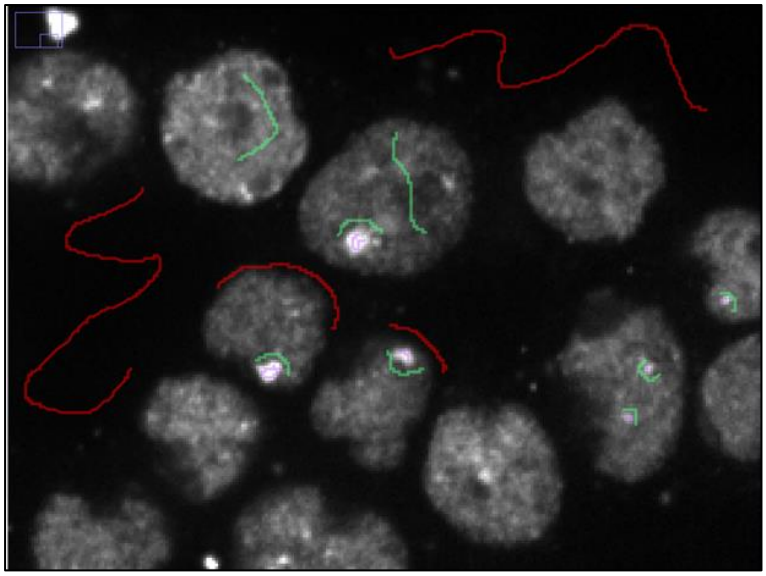

Figure 2. 5 Image showing the learning process of the Trainable Weka Segmentation tool to identify nSBs (purple lines), noise background (green lines) and empty background (red lines).

2.13 Real-Time qPCR

2.13.1 RNA Extraction

In preparation for RNA extraction, all cells, buffers and tubes, were kept on ice to ensure minimal degradation. For this protocol, ReliaPrep™ RNA Cell Miniprep System was used according to the manufacturer's instruction (Promega, UK). Cells were washed in cold PBS, and RNALater solution (Thermo Fisher) was added to preserve RNA until the complete set of plates were ready for RNA harvesting, and plates were kept on Ice. When ready, the RNALater solution was removed, and BL Buffer supplemented with Thioglycerol (TG) was added to lyse the cells. Cells were rinsed off several times and transferred into Eppendorf tubes. 100% ethanol was added to precipitate RNA before loading the aliquots into the mini columns coupled with the collection tube. After brief centrifugation at room temperature, the columns were washed with RNA Wash Solution, and DNase digestion set was added to allow DNase digestion to take place, followed by a 15 min incubation period at room temperature. Series of washing steps were performed by adding Column Wash Solution and RNA Wash Solution in alternation with short centrifugations before eluting the RNA in 50 µL of purified RNAse free water. Aliquots

were kept on ice until the quantification was done, which was performed using the NanoPhotometer™ spectrometer (Implen, Pearl Edition). The value of the concentration was recorded, along with the absorbance ratio values; absorbance ratios of 260nm/280nm should equate to 2.0 +/- 0.05, which indicates a pure RNA sample, whereas the ratio 260nm/230nm should be between 2.0 and 2.2, indicating pure nucleic acid. If this ratio value deviates from this, it indicates there may be protein contaminants within the sample. Therefore, samples outside of these ranges were disposed of. Once all measurements were taken, aliquots were frozen at -20°C for further processing.

2.13.2 cDNA synthesis

RNA was reverse transcribed using Life Technologies Superscript III First-Strand Synthesis System kit according to the manual suggested by the manufacturer (Thermo Fisher Scientific, UK). Reverse transcription of SAFB1, SAFB2, SLTM and HSF1 was carried out with random hexamers, whereas the G-rich SatIII RNAs reverse transcription was carried out using the RSM13 primers (Figure 2. 6).

Random hexamers (50ng/µl) or specific primers (50 µM), dNTP mix (10mM) was added to the 500ng of RNA. These mixtures were incubated at 65°C and then placed on ice. For cDNA synthesis premix, 10x RT buffer, 25mM MgCl2, 0.1M DTT, 40 U/µL RNaseOUT, and 200 U/µL Superscript III RT were brought together. Samples and premixes were added together and incubated at 25°C for 10 minutes, 50°C for 50 minutes and 85°C for 5 minutes using a PCR machine (Techne, UK). RNase H (Thermo Fisher Scientific, UK) was added to the mixtures and left incubated at 37°C for 20 minutes. cDNA samples were stored at −20°C for later use.

2.13.3 RT-qPCR

Following cDNA synthesis, 4µL of each sample was added to the 2x SYBR Premix Green PCR mix (Applied Biosystem, UK) along with forward and reversed primers (2 uM). The SYBR Green dye fluoresces when bound to double-stranded DNA that results from amplification of template cDNA using selective primers. GAPDH was used as an endogenous control to normalise cycle threshold (Ct) values. For each cDNA target, the template from cDNA conversion step was mixed with SYBR Green PCR Master Mix (Applied Biosystem, UK) with specific primers (Sigma-Aldrich, UK). Samples with premix

were loaded into a 96-well plate and detected by Step-One Plus Real-Time PCR Systems (Applied Biosystems). Statistical analysis was carried out using the Δ ΔCT method, in which the housekeeping gene and control conditions were used to calculate the relative expression of the genes of interest.

Primers List:	Human	Application:	Strand:
SAFB1_F	5'-GAGGGACCGAACGGACTGTAG-3'	qPCR	
SAFB1_R	5'-TGGCTGATTTGCGATCCTG-3'	qPCR	
SAFB2_F	5'-GAGCGGACGGTC GTGATG-3'	qPCR	
SAFB2_R	5'-ACTTGCGATCCTGACTCTTGG-3'	qPCR	
SLTM_F	5'-TCTTGATTCTCCAATTCACAGTCC-3'	gPCR	
SLTM_R	5'-TCAACTGATACTCCAAACAAGAAACC-3'	qPCR	
GAPDH_F	5'-TGCACCACCAACTGCTTAGC-3'	qPCR	
GAPDH_R	5'-GGCATGGACTGTGGTCATGAG-3'	qPCR	
SATIII direct	5'-GGAATGGCATGGATTGGAAT-3'	RNA-FISH	C-Rich
SATIII Indirect	5'-ATTCCAATCCATGCCATTCC-3'	RNA-FISH	G-rich
pBR3222	5'-CCGGGAAGCTAGAGTAAGTAG-3'	RNA-FISH	Control
M13	5'-CCGTAAAACGACGGCCAG-3'		
RSM13	5'-CCGTAAAACGACGGCCAGTTCCCTTCCA TTCCATTATTATCC-3'	Reverse Transcription	G-Rich
Hur-98 R	5'-AATCAACCCGAGTGCAATCG-3'	qPCR	G-Rich

Figure 2. 6 List of primers used at the RT-qPCR and RNA-FISH Technique.

2.14 Inhibitors

a) Adenosine, dialdehyde – AdOx

Adenosine dialdehyde (AdOx; Sigma Aldrich) is a global methylation inhibitor which promotes S-adenosyl L-homocysteine (SAH) accumulation through S-adenosyl-L-homocysteine hydrolase enzyme impairment. SAH is a major inhibitor of cellular methyl transferases and prevents methylation of new epitopes [196, 197]. AdOx has been widely used in global methylation inhibition assays since 1974 [196-199] and most of the literature reports show an effective final in vitro concentration of 20 μM of AdOx for periods of time ranging between 24-72 hrs [107, 199]. For the purposes of this study SH-SY5Y cells were seeded in appropriate densities, in either in 24 or 6 well plates 24 hrs prior to treatment. In cell imaging assays, cells were seeded on coverslips as described previously. AdOx was diluted in appropriate volume of fresh media to a final concentration of 20 μM. The cells had then their media replaced with the AdOx containing media and

were incubated for 48 hrs, unless otherwise stated, at 37°C/5% CO2 atmosphere. Pilot experiments have shown that extended exposure to AdOx (72 hrs and beyond) is too damaging to the cells, causing higher death rates. AdOx is commercialised by Sigma in a powder form, which requires solubilisation in oil solvents, therefore DMSO were used as a vehicle and control experiments were carried out adding equal volumes DMSO in the control sample media.

b) PRMT5 Inhibitor: GSK591

The Protein Arginine Methyltransferase 5 (PRMT5) is the major type II PRMT enzyme which catalyses the addition of a second methyl group to histone and non-histone proteins in a symmetric configuration, forming Symmetric Di-Methyl Arginine (SDMA). The newly discovered compound GSK591 has been described an excellent specific PRMT5:MEP50 complex inhibitor. In this study, cellular PRMT5 inhibition was achieved using GSK591 at a final concentration of 10 μM in cell media for 72 hrs. SH-SY5Y cells were plated at suitable concentrations in the same manner as conducted on the AdOx experiments. Similarly, GSK591 is commercialised in a powder form, which requires solubilisation in oil solvents, therefore DMSO were used as a vehicle and control experiments were carried out adding equal volumes DMSO in the control sample media.

Following AdOx and GSK591 treatments, cells were submitted to heat shock induction, as described before and forwarded to immunocytochemistry for further image acquiring and analyses.

2.15 SAFB1 Deletion Mutants:

SAFB1 deletions mutants used in Chapter 3 were kind gifts from Professor Mathias Altmeyer, University of Zurich, Germany (as published in Altmeyer et al. 2013 [11]). Mutated SAFB1 mRNA sequences were inserted in to a pAcGFP-C1 (Clontech) plasmid backbone, under a CMV promoter contained enhanced green fluorescent protein (EGFP) located upstream of SAFB1 sequence [11]. These mutants possessed different domain deletions (Δ) of: ΔSAP (lacking amino acid 1-65), ΔHS (lacking amino acid 210-408), ΔRRM (lacking amino acid 406-484), ΔGlycine rich (amino acid 785-917) and ΔC-terminal domain (amino acid 634-917) (Figure 2. 7). Following amplifications and purification, the mutant constructs were overexpressed in HeLa cells as described in

section 2.5, on coverslips, submitted to HS induction, fixed, and stained for HSF1. Further details on image analyses will be found on the correspondent results section.

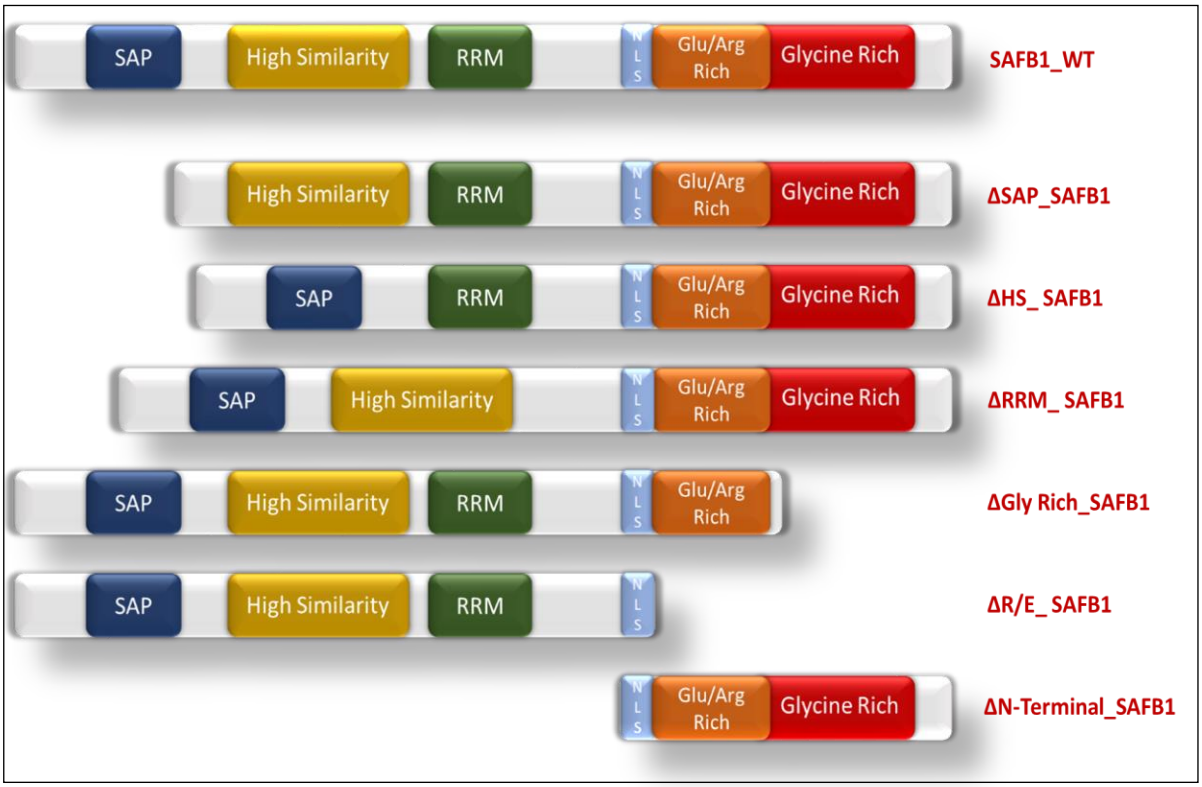

Figure 2. 7 Schematic representation of SAFB1 Deletion Mutants.

2.16 Generation of SAFB2 impaired methylation sites:

The Quick-change II Site-Directed Mutagenesis Kit (Agilent, Technologies, UK) was used to introduce point mutations at RG and RGG sites to switch the arginine on these sites to lysine using a pEGFP.SAFB2 plasmid as a template (Addgene pEGFP backbone). The final construct would then display KG and KGG motif instead. The kit requires the design of specific primers to each point mutation for subsequently cloning. Primers design were carried out as instructed by the kit manufacturer and all the cloning steps.

a) Identifying RG/RGG motifs and primer design:

We used the canonical mRNA sequence for Human SAFB2 (NM_014649.3) from the Gene Bank of the NCBI database to identify the potential Arginine Methylation sites on the protein sequence. Due to the preference of PRMTs for Glycine flanked Arginine, we

mapped all RG and RGG motifs. The Serial Cloner software was used to plot the mRNA sequence, translate it into protein sequence and identify arginine followed by glycine's (Figure 2. 8).

In order to prevent Arginine methylation, we aimed to replace all the RG/RGG sites with KG/KGG sites. Arginine and Lysine have similar changes. Therefore, the conformational distortion would be minimised. Although lysines are also Methylated, the enzymes required for this reaction are the Lysine Methyltransferases (KMT) which differ from the PRMT in their main substrate.

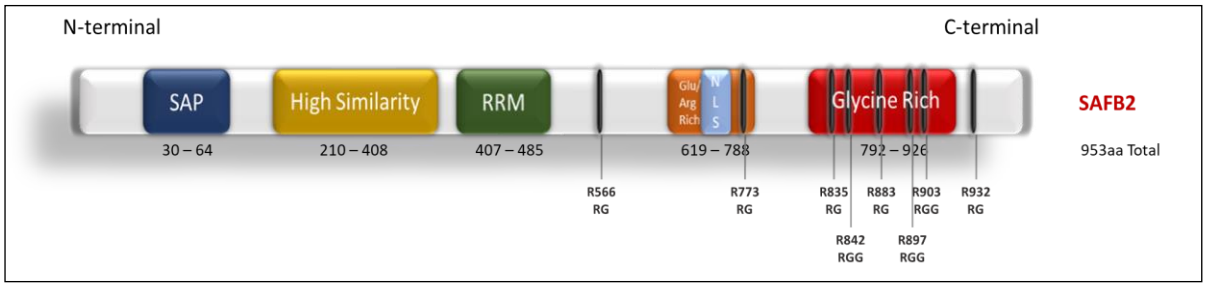

Figure 2. 8 Image of SAFB2 RG/RGG Motifs.

b) Mutating RG/RGG and SUMO sites on SAFB proteins

As we are specifically looking at Arginine methylation, this possibility does not limit our approach. The SUMOylation sites on SAFB1 and SAFB2 are addressed to Lysine 231 and 294. Therefore, the reverse logic was applied, and both K231 and K294 sites were shifted to R231 and R294.

Once the RG/RGG motifs were addressed on the mRNA sequences, point mutations were introduced to shift these arginines to lysines (or vice-versa). Based on these new sequences, primers were designed for each mutation, following the kit manufacturer instructions (Figure 2. 9). The selection of the primers considered optimal length, minimum base-pair alteration needed and melting temperatures.

mRNA	AA position	PTM	New residue	5` -3` Primer sequence:	Length (nt)
SAFB2	R566	Meth	K	CCAAATCAGGAAGC**AaA**GGAATGGAGCGGACG	32
	R773	Meth	K	CGATCATCGAGAC**aaG**GGCCAGTACCAGGAC	31
	R835	Meth	K	AGGCTGAGTGAAGGC**aaG**GGGCTGC	25
	R842	Meth	K	CCTCCCCCC**aaG**GGTGGCCGTGACTGG	27
	R883	Meth	K	CAAGGTGGCGAG**AaG**GGCCTGTCTGGG	27
	R897	Meth	K	CATGGCAAGC**aag**GGTGGAGTGGCGGGG**aaA**GGCGGCTTTG	41
	R903	Meth	K		
	R932	Meth	K	GTGGCCAGCCAGGAC**aaG**GGCAGCAGAGTCCC	32

Figure 2. 9 List of designed primers for SAFB2 mutations introduction.

All primers were ordered from Sigma Aldrich. The mutations were introduced to the template EGFP_SAFB2 bacterial DNA vectors used in our lab, one mutation at a time until all mutations were successfully inserted (aa566, aa773, aa835, aa842, aa883, aa897, aa903 and aa932), in subsequently cloning reactions. The reaction mix for the first mutation (aa566) was prepared, containing 1x reaction buffer, 10ng plasmid template (pEGFP.SAFB2), 125ng forward and reverse primers, dNTPs mix, quick solution and Pfu HF polymerase (50µl total volume). The cycling parameters are shown in Figure 2. 10.

Segment	Cycles	Temperature	Time
1	1		95°C 1
2	18	95°C	50 seconds
		60°C	50 seconds
		68°C	1 minute/kb of plasmid length
3	1	68°C	7 minutes

Figure 2. 10 Cycling parameters used for the PCR reaction.

Following cycling, the reaction tubes were immediately placed on ice for 2 minutes. One microlitre of the Dpn I restriction enzyme was added. After mixing, the reactions were incubated at 37°C for 1 hour to digest the non-mutated plasmid. Two microlitres of the β-

ME mix were mixed with 45µl of XL10-Gold ultracompetent cells. The mix was incubated on ice for 10 minutes with gentle swirling every 2 minutes. Two microlitres of the Dpn I-treated plasmid was added to ultracompetent cells and incubated on ice for 30 minutes. The tubes were heat-shocked at 42°C for 30 seconds and then incubated on ice for 2 minutes. A volume of 0.5ml of preheated (42°C) NZ amine (casein hydrolysate) and yeast (NZY+) broth was added to each tube and incubated at 37°C for an hour. A total of 250µl of the transformation reaction was plated on LB agar plates containing 80µg/ml X-gal and 20 mM Isopropyl-1-thio-β-D-galactopyranoside (IPTG) and kanamycin antibiotic. The plates were incubated at 37°C for >16 hours. After each mutation was successfully cloned into the sequence, the plasmids were amplified, DNA was purified, digested, and run in an agarose gel for band size confirmation. The most promising colonies were sent to sequencing before moving forward to the next site.

c) Plasmid Amplification/Purification - Miniprep

Single colonies were chosen for amplification in 5 – 10 mL of LB media supplemented with Kanamycin (50µg/ml) antibiotic overnight in a shaking incubator at 37°C. 5 mL of bacterial culture was purified on a miniprep scale to isolate plasmid DNA using QIAprep Spin Miniprep Kit (Qiagen, Manchester, UK) according to the manufacturer's instructions.

Bacterial culture was centrifuged for 5 minutes at 10,000 x g at room temperature, and the supernatant was discarded and resuspended in 250µl of purified water. 250 µl of Cell Lysis Solution was mixed with the bacterial suspension and 10 µl of Alkaline Protease Solution. After 5 minutes of incubation, neutralisation Solution (350 µL) was mixed to the lysate and forwarded to centrifugation at 14,000 × g for 10 minutes at room temperature. The supernatant was transferred to a mini-column in a collection tube and centrifuged for 1 minute at full speed. Subsequently, steps involved wash the column with Column Wash Solution twice. Finally, the mini-columns were transferred to a clean 1.5 mL tube, and the DNA was eluted in 50 µl of Nuclease-Free Water and centrifuge at maximum speed for 1 minute at room temperature. Samples were kept at -20°C or used immediately for quantification and digestion with restriction enzymes.

d) Restriction Endonuclease reactions

Appropriate restriction enzymes were chosen using Serial Cloner software to check the size of the plasmids. All restriction enzymes and buffers were purchased from (NEB,

Hitchin, UK) or (Roche, Welwyn Garden, UK). All restriction enzyme reactions were carried out in a total volume of 20µl for 1-2 hours according to the manufacturer's instructions.

e) Agarose gel electrophoresis and sequencing

One percent agarose gels were prepared using agarose (Melford, Suffolk, UK) and 100ml of TAE (Tris / acetic acid / EDTA) buffer with 10 µl of Gel Green dye (Biotium). Digested plasmid DNA samples were mixed with 0.17 volumes of 6 X loading buffer and run on the gel in TAE buffer at 100-110 V until a successful separation of DNA fragments was achieved. One kilobase (1kb) DNA ladder (NEB) was used to estimate the DNA length. Gels were visualised and photographed using a UV transilluminator (Bioimaging systems). Plasmid DNA was sequenced by Source Bioscience (https://www.sourcebioscience.com/) using Sanger sequencing and appropriate primers.

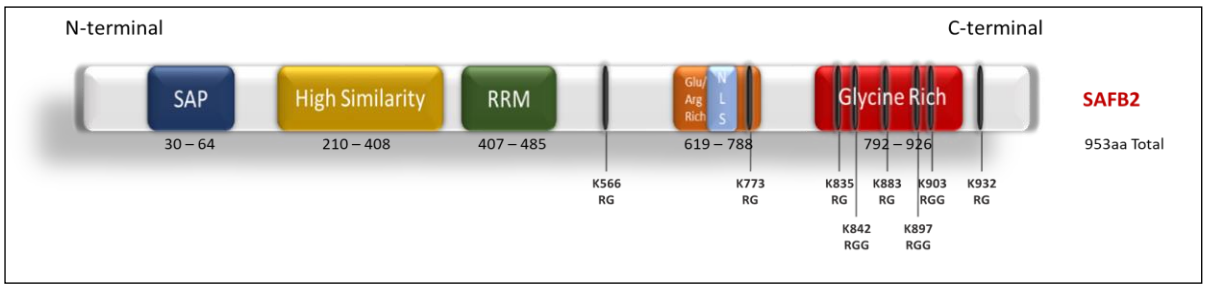

Figure 2. 11 Final Mutated SAFB2_KG/KGG mutated sites.

The final structure of the KG/KGG SAFB2 mutant construct is shown in Figure 2. 11. Following amplification and purification, the mutant constructs were overexpressed in HeLa cells as described in section 2.5, on coverslips, submitted to HS induction, fixed, and stained for HSF1. Further details on image analyses will be found in the correspondent results section.

2.17 SAFB1/SAFB2 Co-Immunoprecipitation

Dr Mamdouh Allahyani conducted Co-IP experiments with my assistance. HeLa cells were seeded in 10cm dishes at 10^5 cells/ml and grown for 48h at 37°C/5% CO_2. Protein lysate was harvested using cold co-IP lysis buffer (section 2.8.4 – RIPA Buffer) with freshly

added protease inhibitors in minimum volume to preserve higher protein concentration. BCA assay was performed to asses protein concentration in the samples and lysate volumes were adjusted to accommodate equal concentrations in all samples.

a) Cross-linking with magnetic beads

Prior to cell seeding, three micrograms of each selected antibody (Rabbit anti-SAFB1 (A300-811A, Bethyl), Rabbit anti-SAFB2 (A301-112A, Bethyl) and Rabbit IgG (2729, cell signalling, USA)) were incubated with 20 µL of Pierce Protein G magnetic beads (Thermo Fisher) and PBS 0.02% tween (PBST) and rotated at 4˚C overnight. Next, beads were applied to a magnet and supernatant was removed. Beads were washed with PBST 3x for 10 seconds with tubes inverted a few times. Beads were rewashed 3x for 10 seconds in coupling buffer (0.2M triethanolamine (Sigma- Aldrich) in PBS with 0.01% Tween-20) at pH9. Next, beads were crosslinked in 20mM Dimethyl pimelimidate dihydrochloride (DMP, Sigma-Aldrich) in a coupling buffer for 1 hour at room temperature. The crosslinked antibodies were quenched in 50mM ethanolamine (Sigma-Aldrich) in PBST for 30min at room temperature. After quenching, beads were washed in elution buffer (0.2M glycine (Sigma-Aldrich) pH2.5 with PBST with vortexing to remove uncross-linked antibody. The crosslinked beads (CLBs) were resuspended in PBST and stored at 4^0C until use.

b) Co-IP

The cross-linked beads were washed 3x in IP buffer before the incubation with cell lysates. Rabbit IgG CLBs were incubated with all the protein lysates for 6 hours at 4°C to remove background interacting events (Figure 2.12). After incubation, beads were applied to a magnet, and the supernatant was collected (pre-cleared lysate). A total of 500µg pre-cleared lysate was incubated with 3µg of each cross-linked antibody (anti-SAFB1, anti-SAFB2 and anti-IgG) overnight at 4°C. The next day, beads were applied to a magnet, and the supernatant was removed and kept (immunodepleted lysate). Beads+linked protein complexes were washed 6 x with 500µL co-IP lysis buffer with gentle mixing. Beads were always kept on ice between washes. After the final wash, 2/3 of the volume of beads+linked protein complexes was stored at -80°C for proteomic analyses. The remainder was resuspended in 2 X SDS gel buffer and Dithiothreitol (DTT, 0.5M), boiled at 100°C for 5 minutes and analysed via western blot as described before.

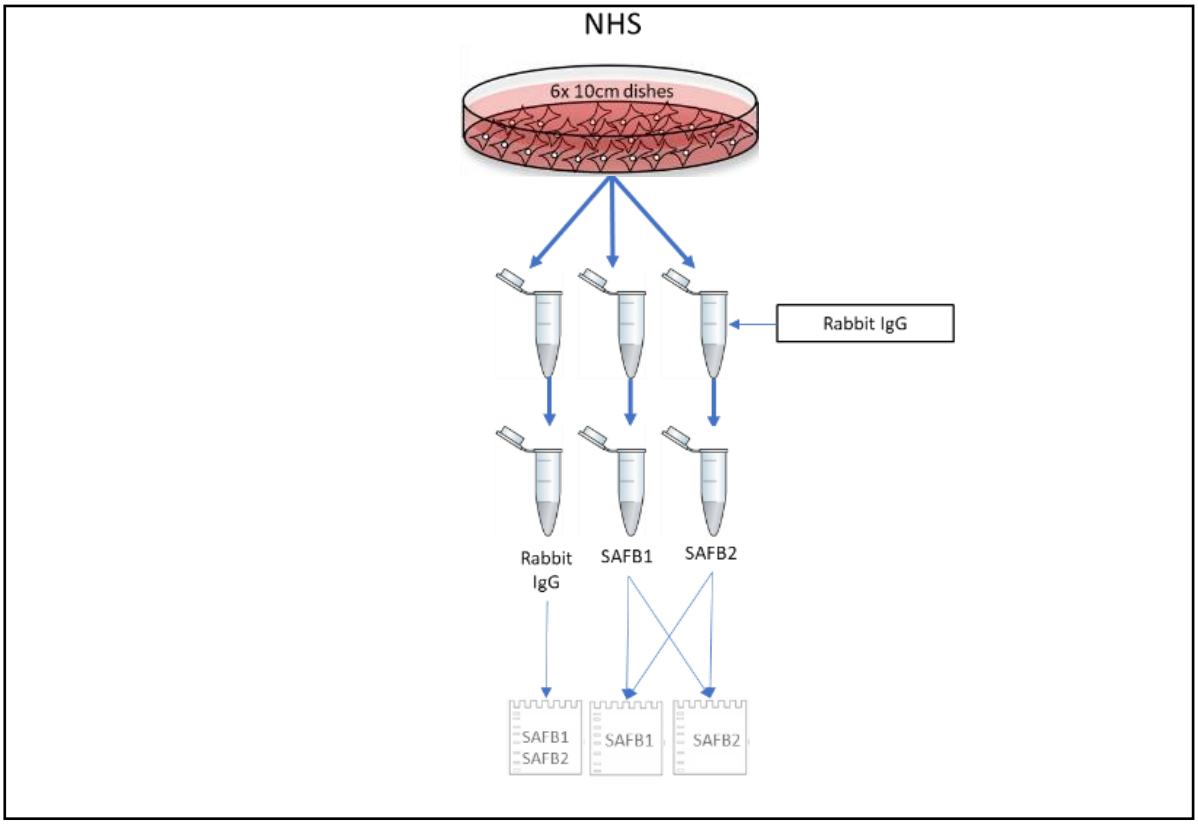

Figure 2. 12 Cycling parameters used for PCR reactions.

2.18 Tandem Mass Tag (TMT) labelling

Preparation and analyses by MS of the SAFB1, SAFB2 and IgG co-IP samples were conducted by Dr Kate Heason at the Proteomic Facility from the University of Bristol. Immunoprecipitated samples were digested on the beads with trypsin and labelled with TMT ten-plex reagents according to the manufacturer's instructions (Thermo Fisher Scientific), then pooled. Samples were then fractionated by high pH reversed-phase chromatography using an Ultimate 3000 liquid chromatography system (Thermo Fisher Scientific). Samples were analysed by Nano-LC Mass Spectrometry.

2.19 2.15.2 Data analyses

The raw data files were quantified using Proteome Discoverer software v1.4 (Thermo Fisher Scientific) and searched against the UniProt Human database. The abundance of proteins was filtered using a 5% false discovery rate (FDR). Data were log2 transformed, so they would be normally distributed. The fold change for SAFB1 and SAFB2 against the

corresponding IgG for each biological replicate was calculated, and the fold change average was also calculated. Paired T-test was performed. Proteins with <2 unique peptides were removed. Only proteins with statistical significance ($P<0.05$) compared to the corresponding IgG were considered for further analyses.

2.20 Validation of Proteomics Results

In order to validate the proteomics results obtained, the Ago2 protein was chosen as one of significant interest identified by the TMT-MS as an interacting partner of SAFB1 and SAFB2. For this, a different cell line from the first IP was used to add robustness to the analyses. SH-SY5Y cells were seeded in 10cm dishes at $2x10^6$ cells/ml and incubated for 48h. After incubation, cells were forwarded to Heat shock induction. Two timepoints were chosen for Co-IP assays: 50% of the plates were kept under normal conditions until cell lysis (NHS – No Heat Shock Control) and the remaining plates were submitted to one-hour of heat stress, and cells lysis were carried out immediately after the heat withdraw (HSNR- Heat Shock No Recovery). Protein lysate was harvested in cold co-IP lysis buffer with freshly added protease inhibitors in minimum volume to preserve higher protein concentration. Samples were forwarded to protein content measurements as described before. The lysates were then diluted in IP Buffer to achieve a final concentration of 2 µg/µL.

SAFB1, SAFB2 and Ago2 Co-IP

Dynabeads Protein G (Life Technologies) were cross-linked with Mouse Ago2 (FUJIFILM Wako Chemicals U.S.A. Corporation, #018-22021), mouse IgG (14-4714-82, Invitrogen), rabbit IgG (2729, cell signalling, USA), customised Rabbit SAFB1 (1895 – Thermo Fisher) and Rabbit SAFB2 (1898) before the experiments, as described above. Figure 2. 13 shows the experimental design of the co-IP. The cross-linked beads (CLBs) were washed 3x in IP buffer prior to the incubation with cell lysates, and 1 mg of beads cross-linked with each antibody was transferred to a new Eppendorf tube. For the pre-clearing step, a minimum of 800 µg of NHS and HSNR lysates were incubated with Mouse IgG and Rabbit IgG CBLs for 30 minutes at 4°C in constant rotation to remove background interacting events.

Following the pre-clearing step, beads were applied to a magnet, and the supernatant was collected (pre-cleared lysate) and transferred to a new Eppendorf tube containing 1 mg of each desired CBLs. Samples were incubated overnight at 4°C in constant rotation. The next day, beads were applied to a magnet, and the supernatant was removed and kept aside (immunodepleted lysate). The beads were washed 6 x with 500 µL co-IP lysis buffer with gentle mixing. Beads were always kept on ice between washes.

Finally, Beads were resuspended in 2 X SDS gel buffer and Dithiothreitol (DTT, 0.5M), boiled at 100°C for 5 minutes and loaded into 8% SDS gels for subsequent western blot analysis (as described above).

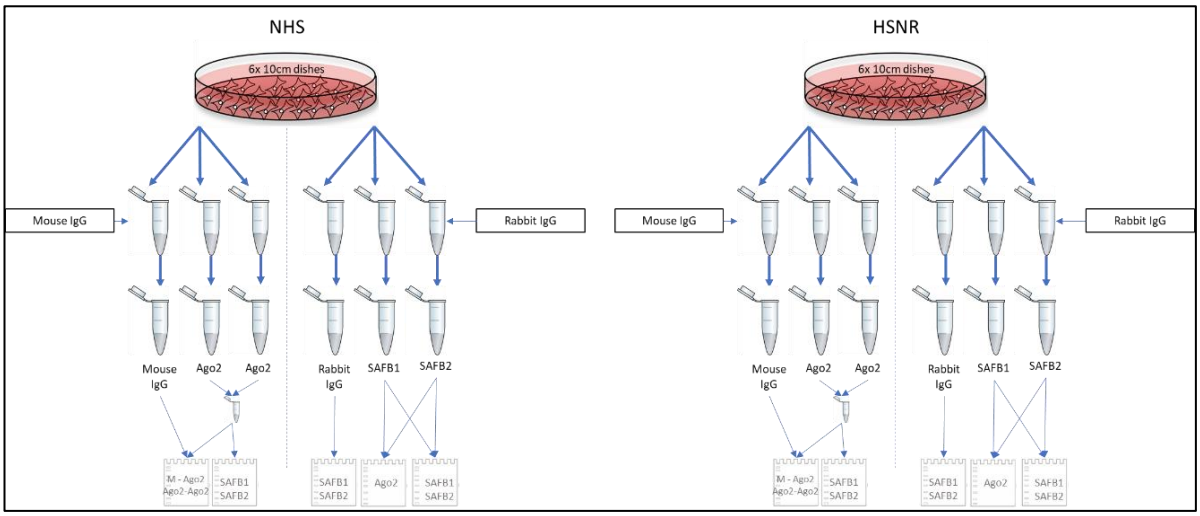

Figure 2. 13 Immunoprecipitation experimental design.

Gel run/analysis

50 µg of protein from the original NHS and HSNR lysates were loaded onto the western blot gels as input. The IP lysates were divided onto 2 lanes to accommodate reverse probing with 2 different antibodies, as described in Figure 2. 14.

IP Lysate	Probed with			
	SAFB1 (Bethyl A300-811A)	SAFB2 (Bethyl A301-112A)	Ago2 (FUJIFILM 018-22021)	SRSF1 (Thermo Fisher 32-6400)
Ago2	X	X	X	
SAFB1 (1895)	X		X	X
SAFB2 (1898)		X	X	X
Mouse IgG			X	
Rabbit IgG	X	X		X

Figure 2. 14 List of antibodies used to probe each IP lysates.

Chapter 03: Characterisation of Nuclear Stress Bodies (nSBs)

3.1. Introduction

Mammalian cells respond to environmental stress by eliciting a Stress Response (SR), which promotes the down regulation of constitutive expression and the transcription of Heat Shock Protein (HSPs) [208-210] which prevent protein aggregation and help refold miss-folded and defective proteins [211]. The Heat Shock Factor 1 (HSF1) transcription factor is the primary coordinator of the SR, and it mediates the transcription of HSP genes and SATIII long non-coding RNAs. Under stress conditions, cytoplasmic HSP90/HSF1-complexes dissociate, and HSF1 becomes phosphorylated, and it trimerises [184]. HSF1 trimers then translocate to the nucleus and bind heat shock elements in the promoter regions of HSP genes. HSF1 also binds an SAT III DNA repeat sequence located in the pericentromeric regions of chromosomes (e.g., 12q on chromosome 9) [41, 178]. Immunostaining of HSF1 following stress shows the nuclear accumulation of the protein in a characteristic pattern called nuclear stress bodies (nSBs). These nSBs are typically comprised of 2 large foci and small HSF1 puncta (or aggregates of these small puncta) scattered throughout the nucleus [200]. HSF1-DNA binding to the 12q9 chromosome recruits RNA polII, leading to the transcription of SATIII sequences into long non-coding RNAs (lncSATIII). lncSATIII transcripts remain tethered and accumulate at the transcription sites, and the repeat sequences bind and sequester RNA processing factors, such as SRSF1, CBP and SAFB proteins, leading to a pan-downregulation of constitutive transcription [106, 179, 190, 191]. SAFB1 is reported to associate with nSBs [1, 11, 74, 106, 191, 202]; however, it is not known if SAFB2 and SLTM also bind with nSBs [2, 212]. In addition, SAFB1 binds DNA and RNA, and it has recently been found to bind major satellite RNA and stabilise heterochromatin. These data suggest that SAFB1 may be involved in SATIII transcription and stabilisation.

According to the Human Protein Atlas (http://www.proteinatlas.org) [213], SAFB1, SAFB2 and SLTM have been identified by immunohistochemistry in 44 human tissues, including high expression in the brain (cortex, cerebellum, hippocampus and caudate), skin and bone marrow [213]. SAFB1 has also been highly expressed in the hippocampus, whereas SAFB2 and SLTM have shown moderate expression levels in the same region [30]. Most studies have used cancer cells such as HeLa cells to study the HSR. However, in 1997 Jolly and collaborators had reported that primary human fibroblasts form HSF1

nSBs upon heat exposure [200]. The authors showed that after one hour of heat stress, 90% of the fibroblasts exhibited large HSF1 containing nSBs. However, the number of HSF1 positive nSBs decreased with the increased cellular passage, suggesting a relationship between HSF1 nSBs formation and senescence.

Chapter Aims:

This chapter aims to characterise nuclear stress bodies using the major markers of the stress response, HSF1, lncSATIII and SRSF1, together with SAFB1, SAFB2 and SLTM in a neuroblastoma cell line and primary human Fibroblasts. The intention was to compare the distribution and recruitment of SAFB1, SAFB2 and SLTM proteins during normoxic, stress and post-stress/recovery conditions. These experiments will determine whether all or only specific SAFB proteins are recruited to nSBs and also allow us to hypothesise as to the consequence of SAFB recruitment on nSB formation and cellular function. Additionally, I knocked down the expression of the specific SAFB proteins and measured lncSATIII expression levels in order to see if there was a relationship between SAFB proteins and the transcription of lncSATIII RNA.

3.2 Results

3.2.1 Time-course of SAFB Nuclear Stress bodies

The recruitment of SAFB to HSF1^{+} nSBs following a heat shock (HS) is a dynamic process and will occur for a a variable time period prior to cessation of the stress response and dissolution of nSBs. Hence, there will be an optimum time point following HS to measure SAFB positive nSBs. It is not feasible to to continuously measure SAFB recruitment in all cells (following a HS) using standard immunocytochemical methods and we therefore used the IncucyteZoon to carry out time-course experiments. These experiments would allow us to determine the time point (post HS) of maximum SAFB protein recruitment and allow us to compare this value with those used in studies of nSBs. The IncucyteZoon system is automated and offers phase contrast and Green/Red fluorescence time-lapse imaging at multiple positions with 4x (0.2 NA), 10x (0.3 NA) or 20x (0.45 NA) lenses. Three EGFP-tagged plasmids encoding SAFB1, SAFB2 and SLTM proteins were transfected into HeLa Cells 48hrs before imaging (methods, section 2.6). Images were taken before and after heat stress induction every 30 minutes for a period of 6 hours. Phase contrast and green fluorescent filter were used for visualisation of the EGFP tag.

A summary of the time-course of SAFB1, SAFB2 and SLTM aggregate formation are shown in Figure 3. 1, Figure 3. 2 and Figure 3. 3. Figure 3. 1 shows the dynamics of SAFB1 aggregate formation before and after heat shock induction. SAFB1 'stress bodies' can be observed after HS induction (1hr HS), and they persist up to 1.5 hrs after cessation of the heat stress. SAFB2 'stress bodies', on the other hand, can be seen immediately after HS induction up to 2.5 hrs after heat withdrawal, as shown in Figure 3. 2, whereas overexpressed SLTM nSBs persist for more than 3.5 hrs of recovery as shown in Figure 3. 3.

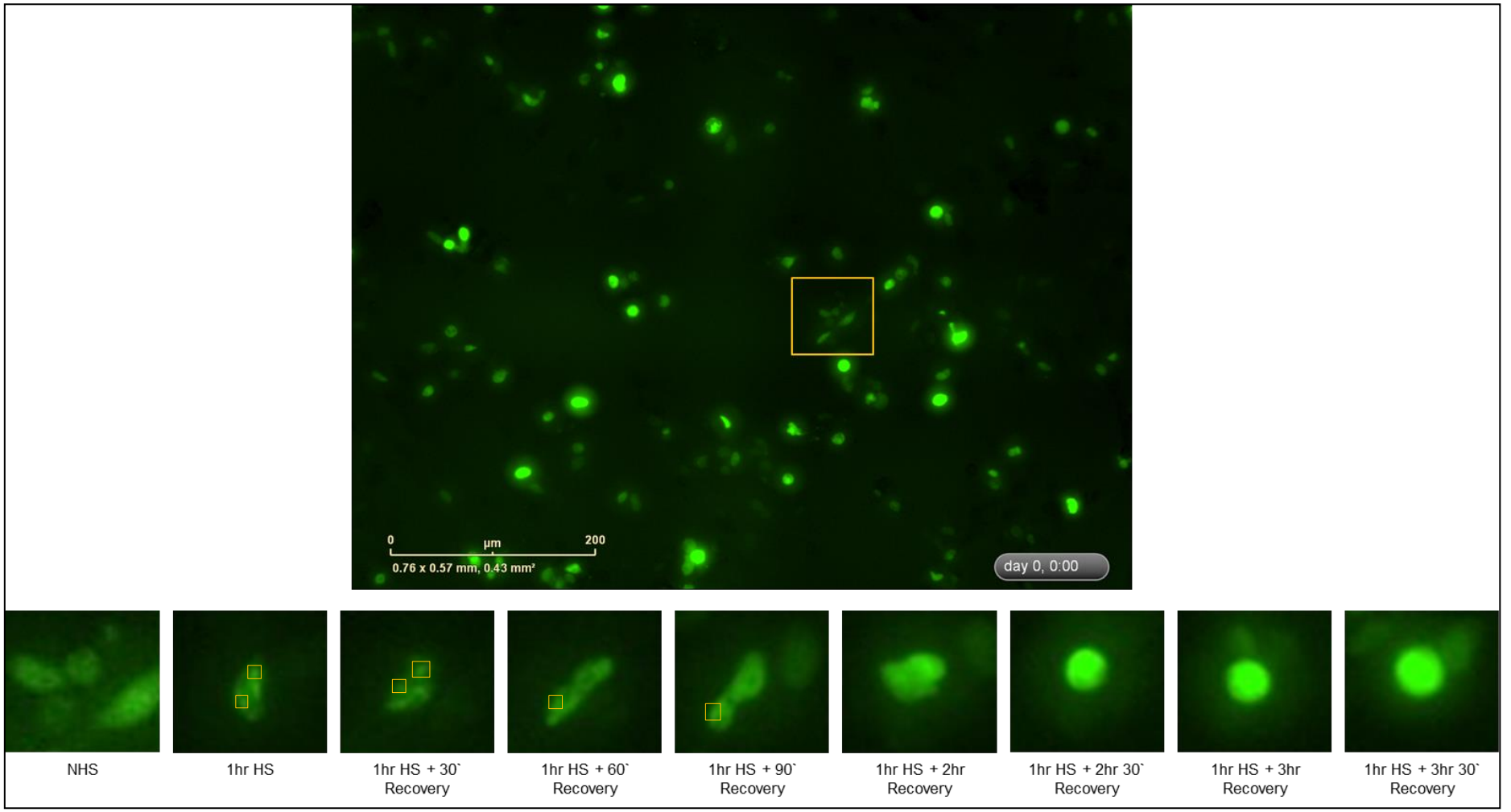

Figure 3. 1 Time course of EGFP-tagged SAFB1 protein overexpression in HeLa cells. Images acquired in the IncuCyte Zoom, before heat stress induction and every 30 minutes following one hour of heat stress treatment up to 3.5 hours. Yellow boxes highlight the formed SAFB1 nSBs.

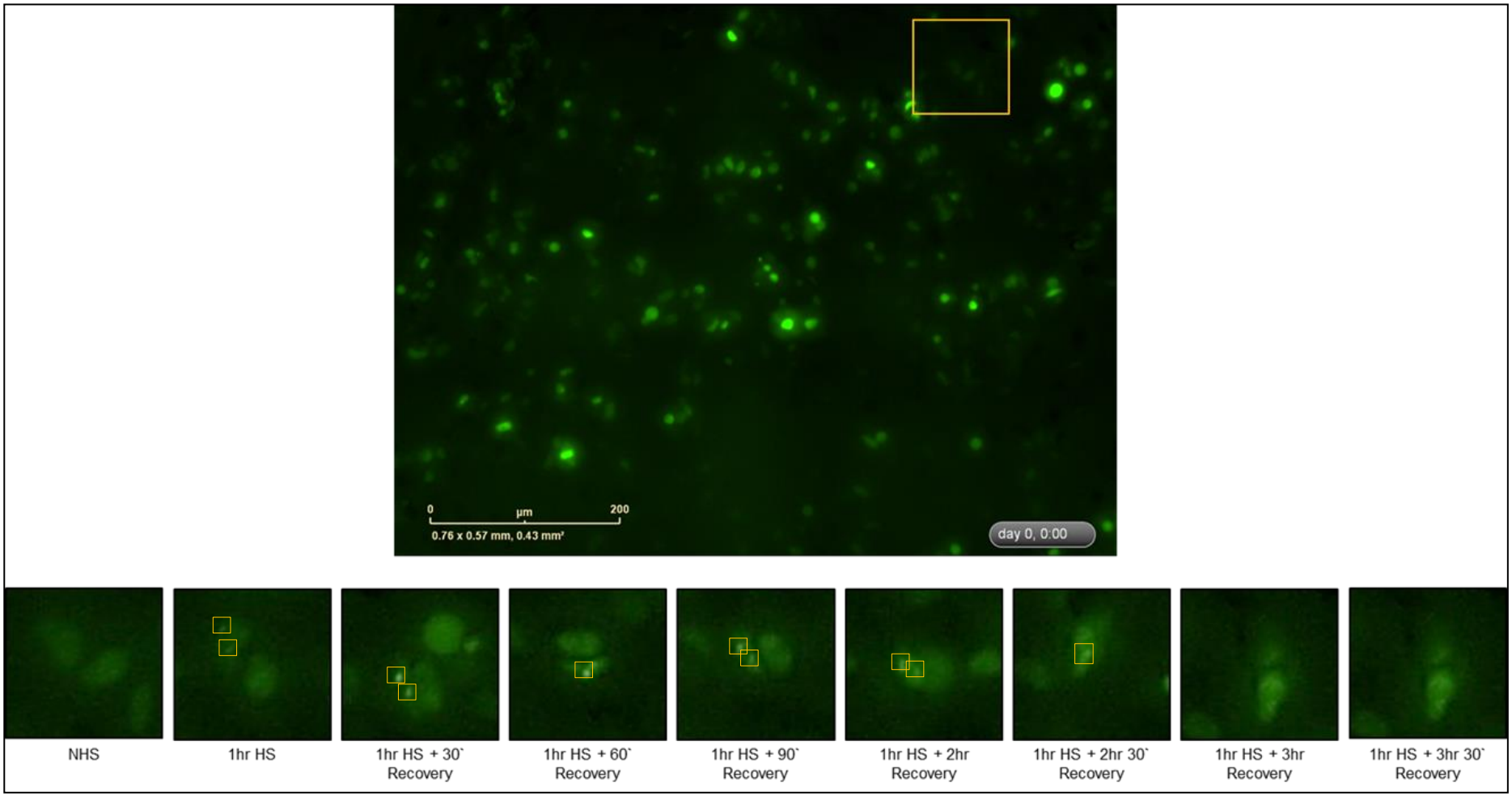

Figure 3. 2 Time course of EGFP-tagged SAFB2 protein overexpression in HeLa cells. Images acquired in the IncuCyte Zoom, before heat stress induction and every 30 minutes following one hour of heat stress treatment up to 3.5 hours. Yellow boxes highlight the formed SAFB2 nSBs.

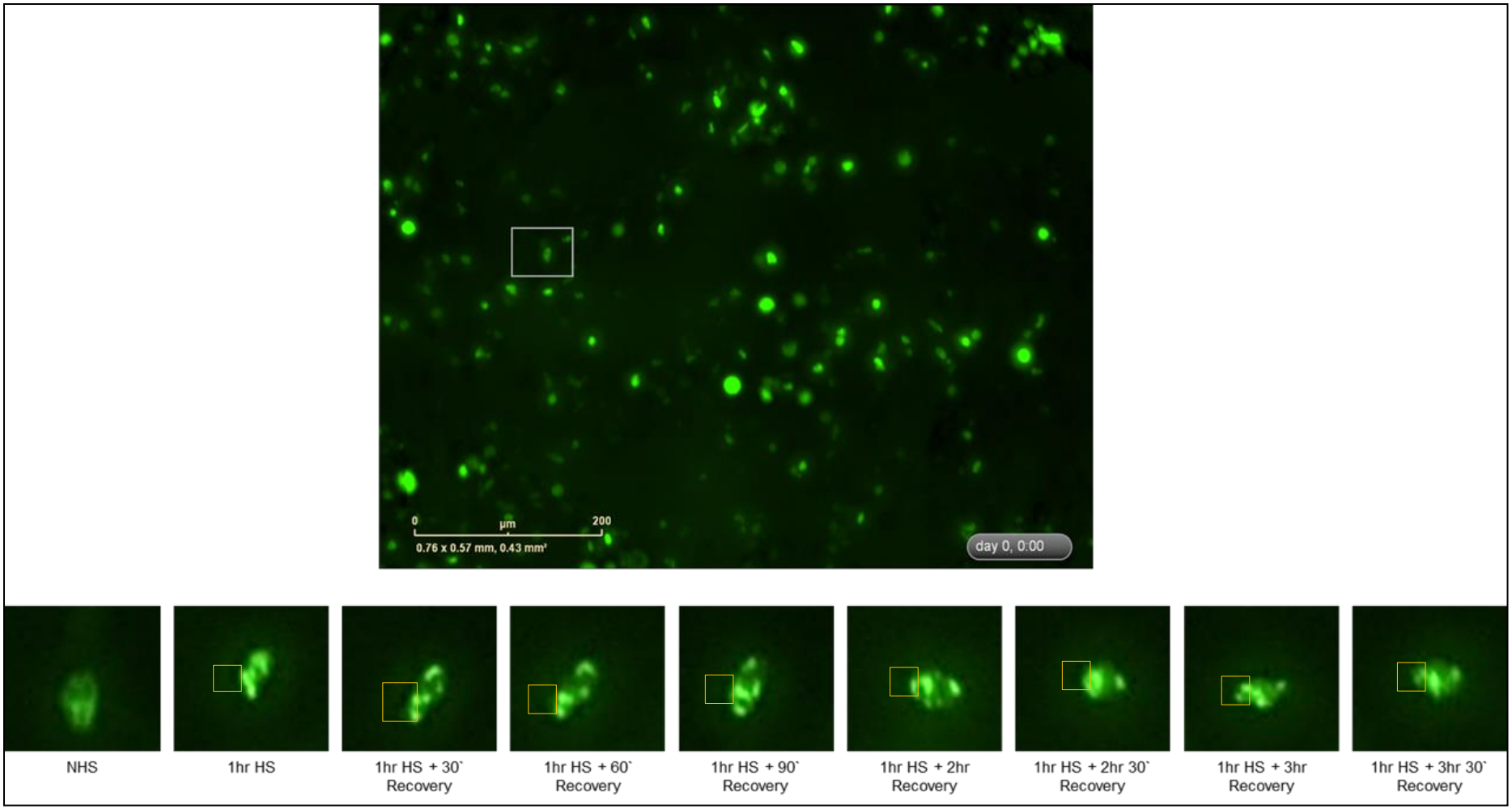

Figure 3. 3 Time course of EGFP-tagged SLTM protein overexpression in HeLa cells. Images acquired in the IncuCyte Zoom, before heat stress induction and every 30 minutes following one hour of heat stress treatment up to 3.5 hours. Yellow boxes highlight the formed SLTM nSBs.

The IncuCyteZoom live Imaging System allowed the visualisation of SAFB protein aggregation/ nsB formation following stress. However, there were some limitations, for instance, the magnification was limited and did not allow high-resolution images to be captured, and only one tagged protein could be visualised, prohibiting co-localisation studies. Additionally, the heat-shock step had to be carried in a separate incubator, and images could not be acquired during the heat shock induction. Nonetheless, the system fulfilled its purpose regarding providing the time-point at which the maximum number of SAFB puncta (nSBs) was formed and this was comparable to that used in similar studies.

3.2.2 Characterisation of SAFB antibodies

Following the time-lapse experiments, it was decided that further characterisation of SAFB protein association with nSBs would be carried out by immuno-labelling followed by conventional light/confocal microscopy. Prior studies had shown SAFB1 was co-localised in nSBs with HSF1, SRSF1 or lncSATIII [1, 11, 74, 78, 191, 202]. These studies used the same anti-SAFB1 antibody (anti-HAP antibody - SAF-B (4D6): sc-135618, Santa Cruz Biotechnology). However, the antibody was raised against antigens that were common to SAFB1 and SAFB2 proteins. In this study, I, therefore, characterised Bethyl Laboratories antibodies raised against aa sequences that were reported to be unique to SAFB1, SAFB2 and SLTM (Figure 3. 4). The SAFB1 antibody (A300-811A) was raised against aas 775 and 825, the SAFB2 Ab (A301-112A) was raised against aas 175 to 225, and the SLTM Ab (A302-835A) was raised against aas 800 to 851. Serial Cloner 2.6.1 software was used to identify the epitopes targeted by these antibodies and were aligned using European Bioinformatics Institute (EMBL-EBI) databases. No alignments longer than two amino acids could be found between the SAFB proteins, suggesting that the antibodies would differentiate between the SAFB family members (Figure 3. 5).

Antibody specificity was also accessed by knocking down the targeted proteins using siRNAs followed by western blotting. SH-SY5Y cells were transfected with 60pM of siRNAs against either SAFB1, SAFB2 or a non-target control sequence. In addition, a fourth group was added, where cells were transfected with both siRNAs (30pM of SAFB1 siRNA and 30pM of SAFB2 siRNA) to promote a double knockdown. The polyvinylidene difluoride (PVDF) membranes were then stained for both SAFB1 and SAFB2 antibodies.

The results showed that the antibodies specifically recognised their target SAFB1, SAFB2 and SLTM epitopes (Figure 3. 6).

Protein/Ab	Manufacturer	#cat	Specificity (aa)	Aa Sequence
SAFB1	Bethyl	A300-811A	775 to 825	REGQHYPERHGGPERHGRDSRDGWGG YGSDKRMSEGRGLPPPPRGRRDWGD
SAFB2	Bethyl	A301-112A	175 to 225	RYPKNFSDSRRNEPPPPRNELRESDRREV RGERDERRTVIIHDRPDITHPR
SLTM	Bethyl	A302-835A	800 and 850	RYPKNFSDSRRNEPPPPRNELRESDRREV RGERDERRTVIIHDRPDITHPR

Figure 3. 4 Bethyl antibodies data about SAFB1 SAFB2 and SLTM antibodies specificity provided by the manufacturer. Adapted from https://www.bethyl.com/.

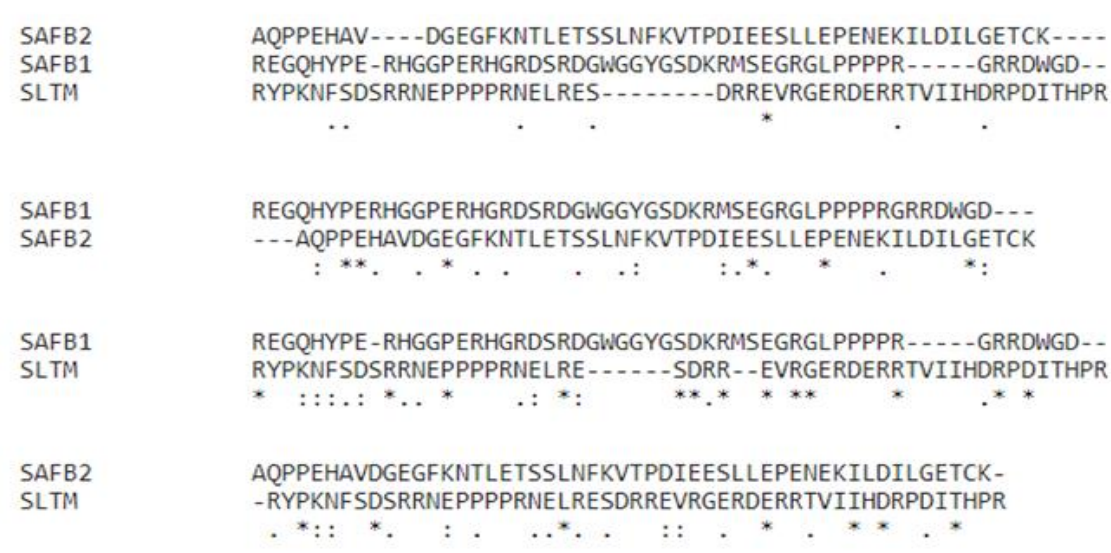

Figure 3. 5 Blast analysis of Bethyl antibodies` recognition sites for SAFB1, SAFB2 and SLTM showing low levels of homology between the recognized motifs. No more than two equal consequent sequential aa are found in homology among the sequences.

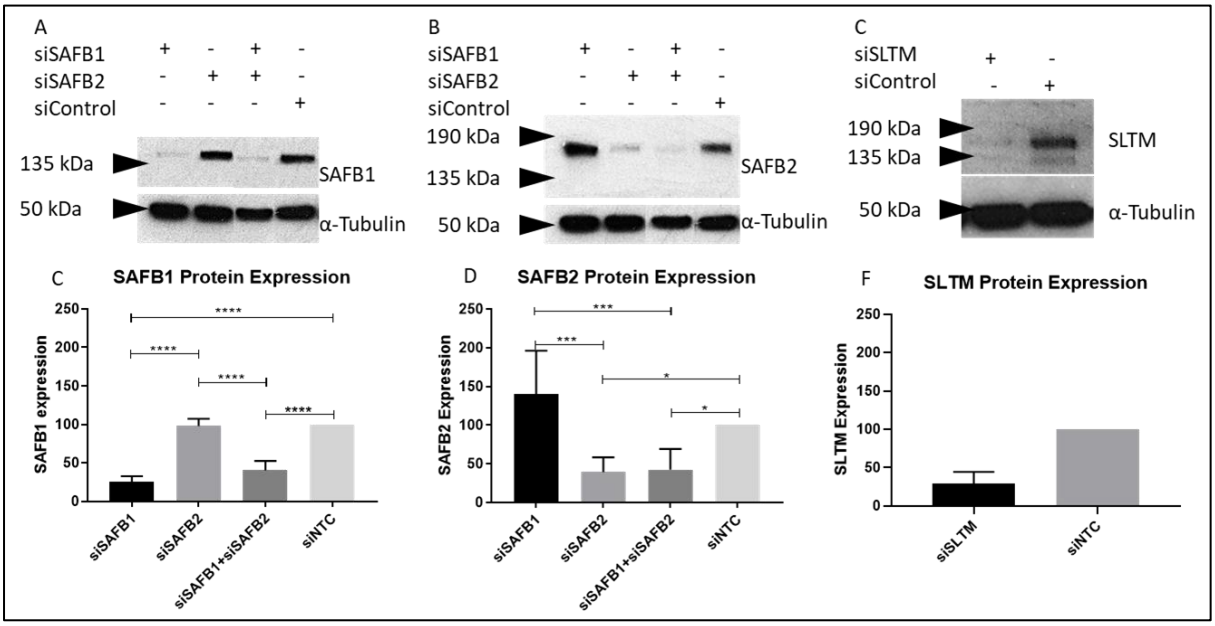

Figure 3. 6 A, B and C: Western Blot, 60 µg of Hela Cells protein lysate probed with specific antibodies against either (A) SAFB1 (A300-811A Bethyl), (B) SAFB2 (A301-112A Bethyl) or (C) SLTM (A302-835A, Bethyl). Cells were treated with 60 pM of siRNA in single or double (30 +30) knockdowns for 48hrs prior to collection. D, E and F: Graphs showing the level of expression of either (D) SAFB1, (E) SAFB2 or (F) SLTM. SAFB1 and SAFB2 values represent the means of three independent experiments, whereas SLTM values represent two independent experiments.

3.2.3 Immuno-Characterization of nSBs:

nSBs are primarily defined by the presence of lncSATIII and co-localisation with HSF1, SAM68 and Serine/arginine-rich splicing factor 1 (SRSF1) and SRSF9 [214]. In this study we therefore used lncSATIII, HSF1 and SRSF1 to characterise nSBs and ensure our results were comparable with previous studies. The following primary antibodies were used: Rat Anti-HSF1 (ADI-SPA-950-F, ENZO Life sciences), Mouse anti-SRSF1 (32-4600, ThermoFisher Scientific) and Fluorescein Avidin D (A-2001, Vector Labs) against biotinylated RNA used in the RNA-FISH protocol. These antibodies are well characterised in the literature and proven to be specific to their targets [215-219]. Secondary antibodies, donkey anti-rabbit and donkey anti-rat (711-165-152, 712-165-150 respectively, Jackson Laboratories) against the primary antibody hosts were conjugated with either Cyanine 2 (Cy2) or Cyanine 3 (Cy3) to provide fluorescence. Dilutions were made in a 1 in 200 ratio in both steps. In addition, cells were stained with Hoechst or DAPI for nuclear visualisation. For the RNA-FISH staining, an indirect biotinylated-RNA probe against the nucleotides 97–106 of the satellite III sequence of human chromosome 9 (accession number: X06137) was used for the hybridisation [202] and an amplification step was added, in which a Biotinylated Anti-Avidin D (BA-0300, Vector Labs) was used to amplify the Fluorescein Avidin D signal. An oligonucleotide complementary to the pBR322 sequence was also used as a non-targeting control.

In the absence of the heat shock, both HSF1 and SRSF1 were distributed evenly throughout the nucleus, showing no distinct pattern, and there was no staining for SATIII transcripts as assessed by RNA-FISH (Figure 3. 7). Following heat stress, HSF1 and SRSF1 aggregates formed, and there were large puncta and/or smaller spherical puncta in the nucleus that persisted for at least one hour after the HS (two hours from the commencement of HS). Interestingly, both HSF1 and SRSF1 formed two large puncta and multiple small foci in the nuclei following HS (Figure 3. 7). These findings are in accordance with Jolly et al. (1999), who showed that when primary human fibroblasts were exposed to a HS formed HSF1 formed aggregates that can be classified in three categories accordingly to their size: (1) two large clusters of granules, (2) highly compact foci and (3) small punctate foci [200]. These two large clusters of HSF1 are from now on referred to as canonical nuclear Stress Bodies (nSBs).

Similarly, the SATIII transcripts which were absent before heat stimuli are now visible in the nucleus in the form of large two large foci and/or a few small foci (Figure 3. 7). The same pattern persists for at least one hour after the heat stress is withdrawn [41, 178].

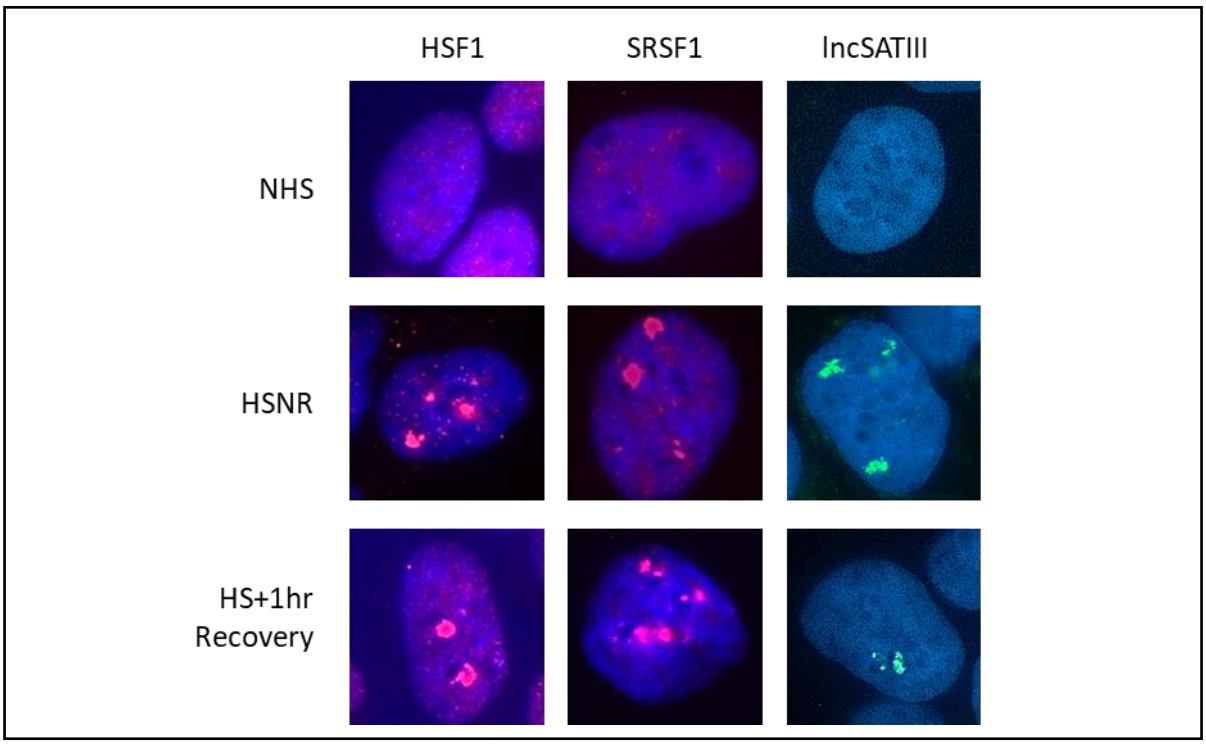

Figure 3. 7 SH-SY5Y Cells immunostained for the major markers of the stress response, HSF1, SRSF1 (red) or SATIII RNA (green). Cells were submitted to either one hour of Heat Shock with No Recovery (HSNR), one hour of Heat Shock followed by 1hr of Recovery (HS+1hr Recovery) or No Heat Shock (NHS – Control) prior to the fixing. Oil 63x objective, 5x zoon.

The results also show that HSF1 foci co-localise with the lncRNA SATIII foci, and this co-localisation persists for at least 1 hour after termination of the HS (Figure 3. 8). This is in accordance with the findings of Jolly, who showed HSF1 binds the SATIII DNA in pericentromeric regions upon heat stress and drives the recruitment of RNA polymerase II, leading to the transcription of the SATIII repeats into lncRNAs [41, 178, 179, 200].

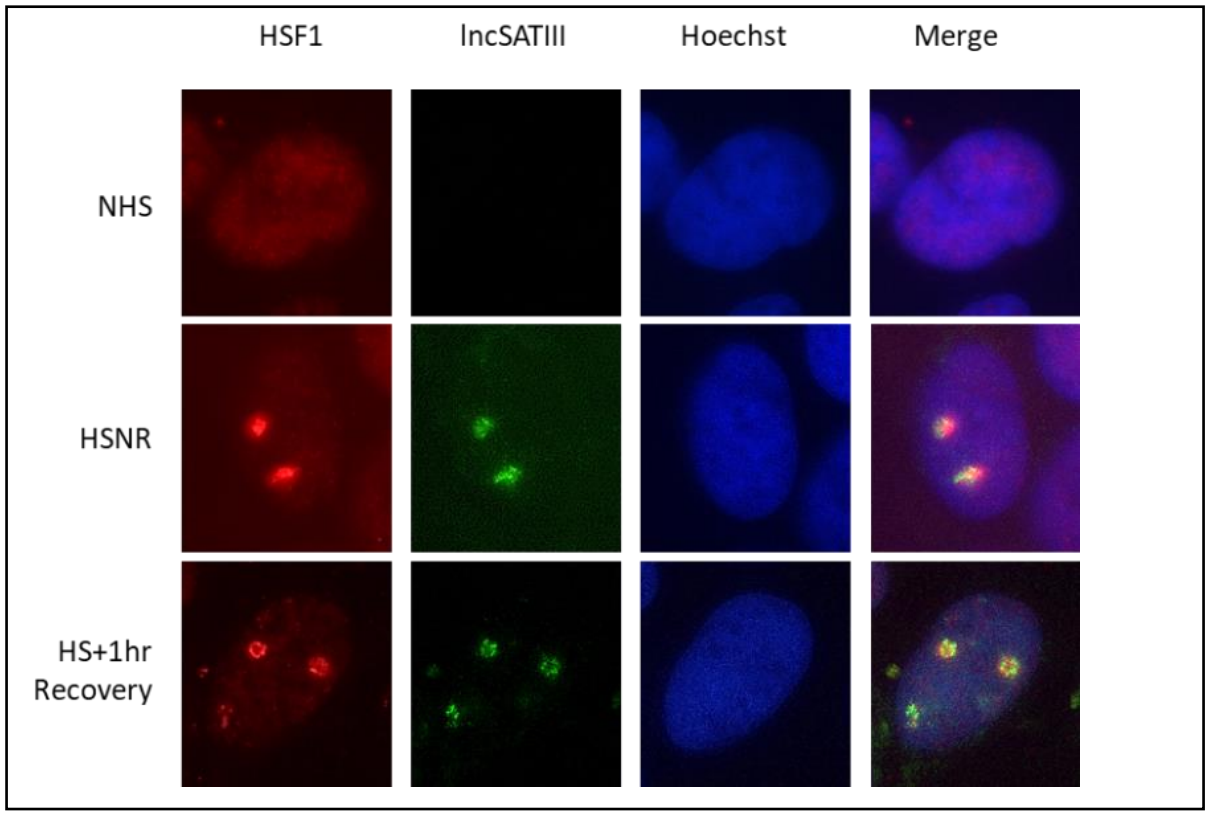

Figure 3. 8 SH-SY5Y cells immunocytochemically stained for HSF1 and stained for SATIII RNA using FISH. Cells were subjected to either No Heat Shock, 1hr of Heat Shock or 1hr of Heat Shock followed by 1hr of Recovery prior to the fixing. Images are using an Oil 63x objective.

3.2.4 Recruitment of SAFB1, SAFB2 and SLTM to HSF1 stress bodies

After ensuring we were producing canonical SBs, as defined previously, we assessed the recruitment of SAFB proteins. Previous studies used an anti-HAP antibody (polyclonal antiserum affinity-purified on recombinant His-tagged HAP [1]) that cross-reacted with SAFB2. Therefore, to investigate the distribution of individual SAFB proteins, HS cells were stained with anti-SAFB1, SAFB2 and SLTM antibodies (Figure 3. 9).

In the absence of stress, SAFB1 displays a diffuse granular staining pattern throughout the nucleus with few aggregation regions, except in the nucleoli. Although the literature has shown SAFB1 as a nuclear protein [1, 6, 7], it is possible to identify minimal cytoplasmatic staining, which may be explained by tissue-function specificity, as previous studies were conducted in different cell types. After heat stress induction, minimal changes in SAFB1 distribution are observed. Very subtle intensification of the SAFB1 aggregation regions can be observed. However, they are difficult to identify under the microscope. Under normal conditions, SAFB2 is also evenly distributed across the nuclear compartment, with a punctum of intense staining. SH-SY5Y cells also display a nuclear SAFB2 distribution with no nucleoli staining [4]. Upon heat exposure, a portion of the cell population forms large sites of SAFB2 aggregation. These newly formed SAFB2 puncta display high-intensity staining and are easily identified under microscope observation (Figure 3. 9, arrow). Finally, SLTM has a diffused granular distribution across both nucleus and cytoplasm, with predominant cytoplasmic staining, and this is expected due to the absence of an NLS and is in accordance with the literature [5, 8]. Heat exposure leads to the formation of SLTM accumulation sites within the nuclear compartment in a small portion of the cell population, suggesting that SLTM is imported to the nucleus under stress conditions. One hypothesis is that SLTM could enter the nucleus via association with SAFB1 or SAFB2 or another protein-binding partner. SAFB1 have been shown to self-associate and directly interact with SAFB2 via the N-terminal region [4]. Since SLTM and SAFB1 N-terminal regions have 44% homology [2], SLTM can possibly directly interact with either SAFB1 or with a SAFB1 biding partner. However, SLTM is yet to be proven to interact with either SAFB1 or SAFB2 under stress conditions.

The immunostaining reveals distinct distribution patterns between SAFB1, SAFB2 and SLTM not shown before (Figure 3. 9). The comparative image shows patterns of

distribution and recruitment of SAFB1, SAFB2 and SLTM, suggesting that they have distinct roles within the cell. This distribution also suggests that the SAFB proteins will bind different proteins via their C-terminal domains.

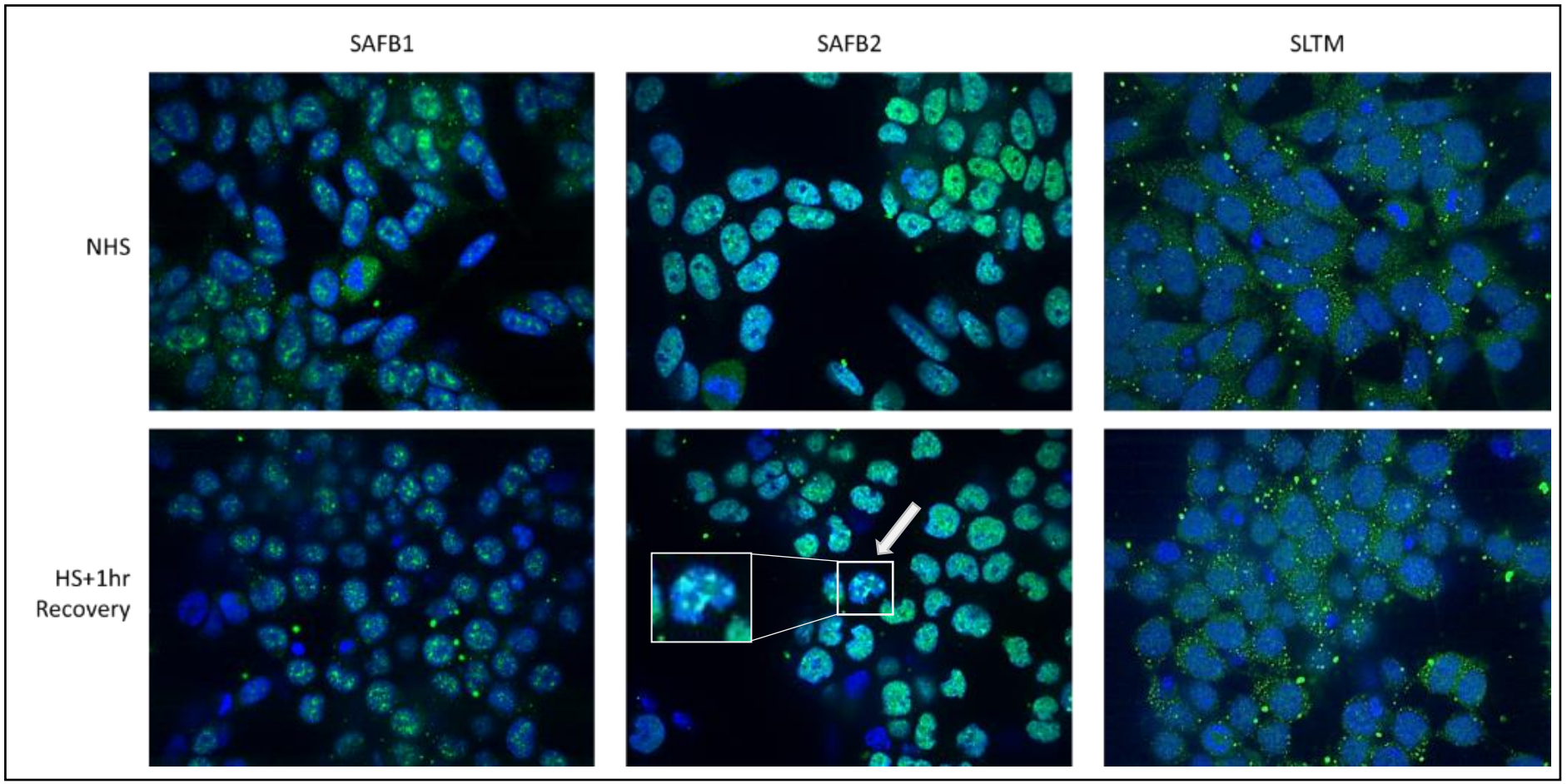

Figure 3. 9 Distribution pattern of SAFB1, SAFB2 and SLTM before and after heat stress. SH-SY5Y cells were submitted to no stress or one hour of heat stress followed by one hour of recovery prior to fixing. Staining for either SAFB1, SAFB2 or SLTM was carried out with specific antibodies. Water 60x objective.

Next, I investigated whether SAFB1, SAFB2 and SLTM co-localised with HSF1- positive nSBs following an HS. Under no heat-shock (NHS) normal conditions, SAFB1 was distributed throughout the nucleus with some small puncta (as described previously in Figure 3. 9), whereas HSF1 was evenly distributed throughout the nucleus, and both SAFB1 and HSF1 were absent from nucleoli. Following an HS, HSF1 staining was again seen in large foci (denoting nSBs) of intense fluorescence easily distinguished by microscopic observation (Figure 3. 10). These HSF1 positive nSBs corresponded to that co-localising with SatIII (Figure 3. 8). However, SAFB1 did not co-localise with HSF1-positive nSBs after HS or after HS plus one-hour recovery (Figure 3. 10 A). This contradicts previous studies that found SAFB1 co-localised with HSF1 positive nSBs following an HS [1, 74, 78, 191, 202], although these studies applied the anti-HAP antibody, which is not specific for solo SAFB1 staining.

Experiments carried out with the anti-SAFB2 specific Ab found that under normal conditions (NHS), SAFB2 was evenly distributed across the nuclear compartment, excluding the nucleoli and showing a few small regions of intense fluorescence aggregation (Figure 3. 10 B). However, SAFB2 co-localised with HSF1-positive nSB structures after HS and following HS plus one-hour recovery (Figure 3. 10 B).

SLTM immunostaining (Figure 3. 10 C) showed a predominantly granular cytoplasmatic staining. Following HS and HS plus recovery, there were some puncta like SLTM expressions within the nuclear compartment that co-localised with HSF1-positive nSBs. This is in accordance with the previous findings [15] that showed SLTM co-localises with a subset of lncSATIII nSBs.

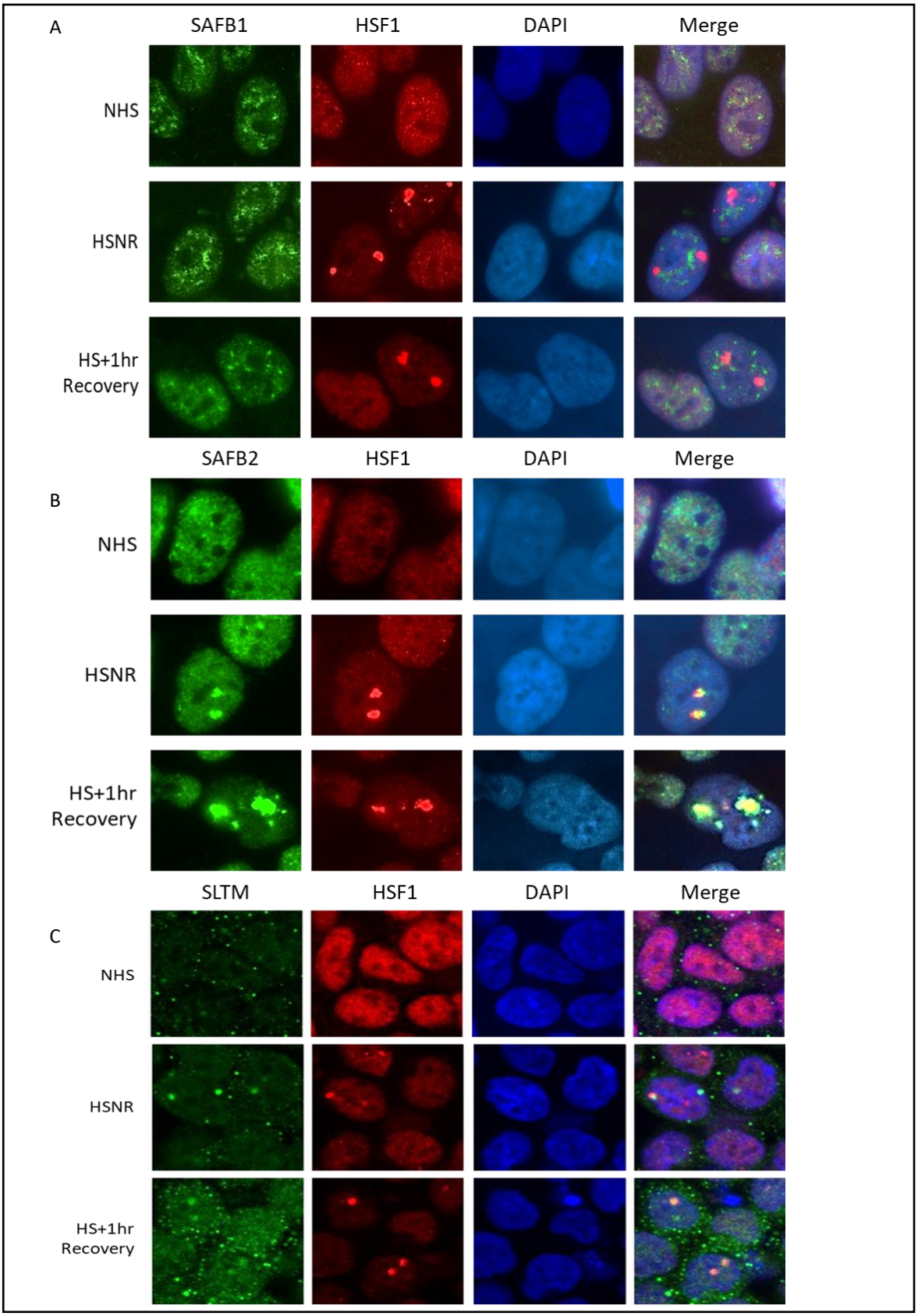

Figure 3. 10 Distribution of either SAFB1 (A), SAFB2 (B) or SLTM (C) and HSF1 before and after heat stress. SH-SY5Y cells were summited one hour of heat stress or one hour of heat stress followed by one hour of recovery prior to fixing. Staining was carried out using specific antibodies.

Next, I evaluated the recruitment of SAFB1 and SAFB2 to SRSF1+ nSBs. SRSF1 has been shown to co-localise with HSF1 positive nSBs [180]. I, therefore, investigated whether SAFB1 or SAFB2 (Figure 3. 11 A and B) colocalize with SRSF1 following an HS. SAFB1 did not co-localise with the large SRSF1 puncta formed following an HS or HS plus recovery (Figure 3. 11 A). However, SAFB2 did co-localise with SRSF1-positive nSBs following HS and HS+R (Figure 3. 11 B).

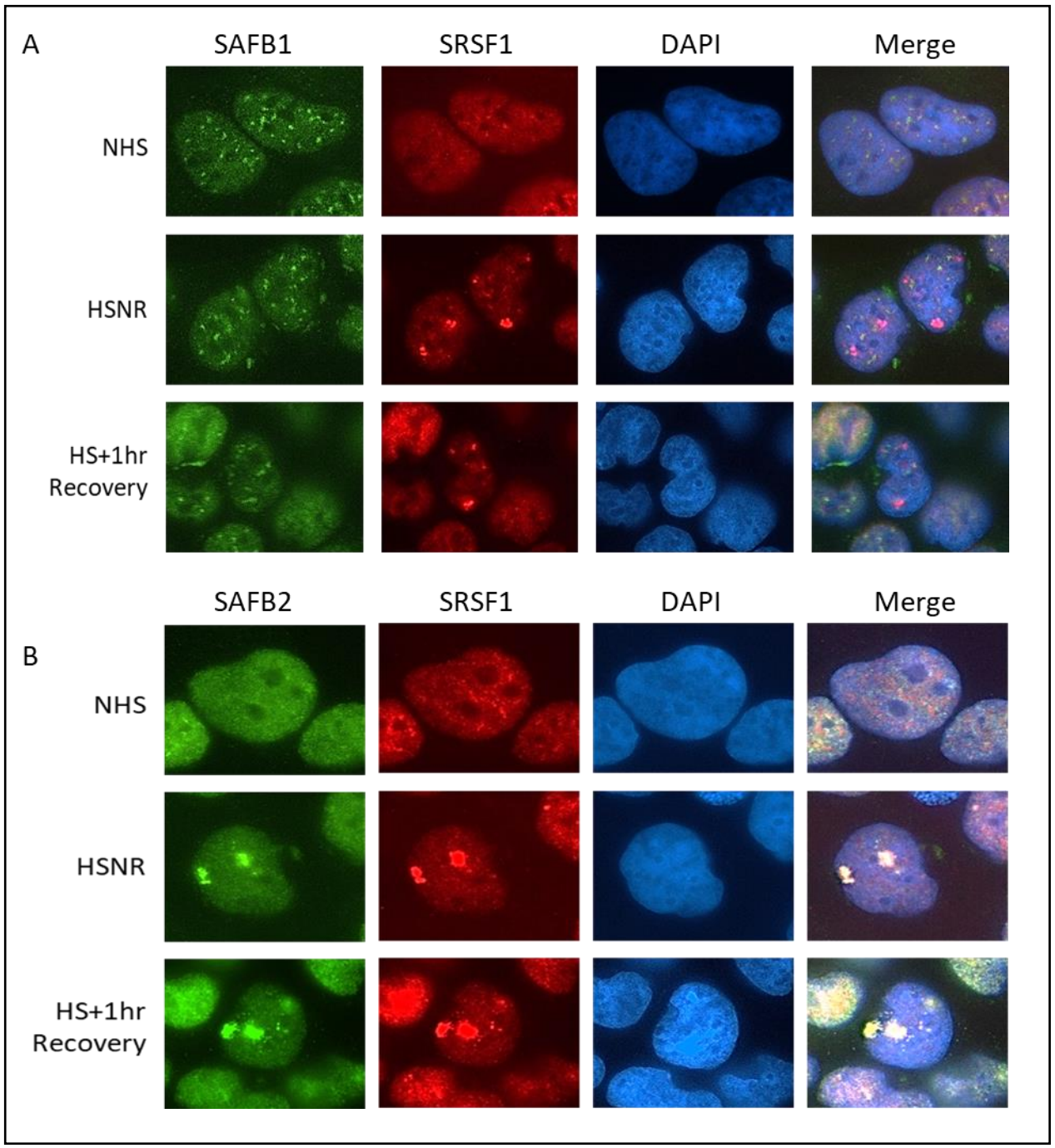

Figure 3. 11 Expression pattern of either SAFB1 or SAFB2 and SRSF1 before and after heat stress. SH-SY5Y cells were submitted to either one hour of HS or one hour of HS followed by one hour of recovery prior to fixing. Hoechst was used for nuclear identification (middle right). Merge images of SAFB1 and SRSF1 shows no co-localization between the proteins (far right). Merge images of SAFB2 and SRSF1 shows an almost complete co-localization between the proteins (far right) in both time points following heat stress Water 60x objective, 5x Zoom.

3.2.5 Development of an Automated nuclear Stress Body`s identification system

In order to characterise and/or quantify SAFB1, SAFB2 and SLTM recruitment to HSF1 positive nSBs in an efficient and unbiased manner, our colleagues from the Wolfson Bioimaging Facility, University of Bristol, developed an automated analysis protocol for the Modular Image Analysis (MIA) plug within the ImageJ/Fiji software. Details on the development of the analysis protocol and optimization can be found in Chapter 2, Methods, section 2.12. After optimising the MIA parameters, output images were accurate enough to identify large and small nSBs stained for SAFB1, SAFB2, SLTM, HSF1 and, further on, lncSATII RNA. Figure 3. 12 shows the output recognising image generated by the MIA.

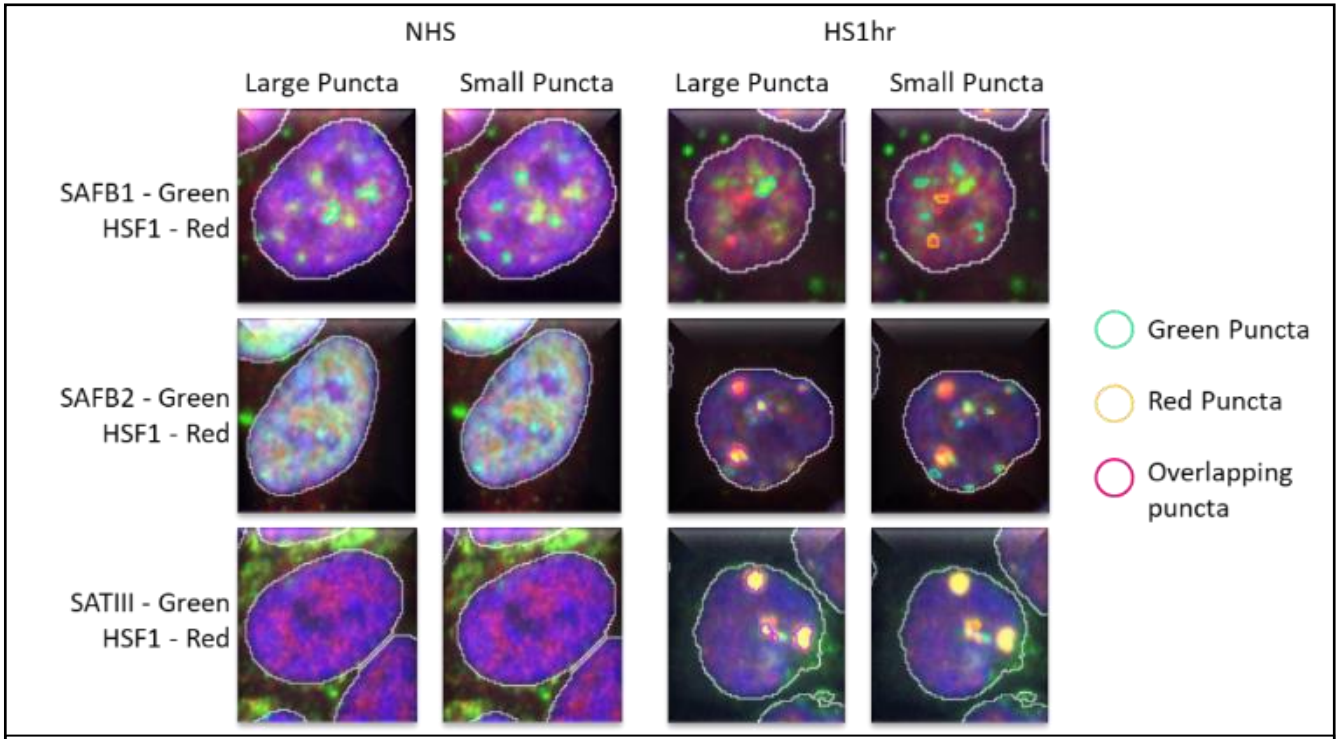

Figure 3. 12 Modular Image Analysis (MID-ImageJ plugin): output from the image analysis software is shown. The puncta identified are highlighted as indicated in the legend. Following identification, puncta are sorted according to size (Small<0.7µm2, Large>0.7µm2<10µm2) and localization (with the nucleus).

3.2.6 MIA Analysis: Formation of nSBs and SAFB proteins recruitment

To characterize and/or quantify the recruitment of SAFB proteins to the HSF1$^+$ nSBs, I induced thermal stress in SH-SHSY cells, fixed them at the previously established timepoints (no HS (NHS), 1hr of HS with no recovery (HSNR) and 1hr HS + 1hr recovery (HS1hr), and stained them with HSF1 and either SAFB1, SAFB2 or SLTM antibodies (methods, sections 2.2-2.4, 2.9). Images were then acquired using the Perkin Elmer Opera LX a high content imaging system. Images were taken using the same parameters for each antibody/condition (methods, section 2.11). Following the successful development of the automated analysis protocol, the acquired images were then loaded into the MIA plug-in in the Fiji/ImageJ software (more details in Methods, section 2.12). The final quantitative results were then expressed as the percentage of cells expressing the antibody-specific to nSBs, relative to the total cell count. Whereas co-localisation was expressed as the percentage of cells expressing co-localising SAFB/HSF1 nSBs within the HSF1$^+$ nSBs cell population.

In order to discriminate between canonical HSF1-positive nSBs and other HSF1 positive foci, the HSF1 puncta were divided into two populations based on their size (Figure 3. 13). In the absence of heat shock (NHS), very few cells had large HSF1 puncta (0.0785% ± 0.02%). Immediately after heat shock (HSNR), there was a significant increase in the proportion of cells expressing large HSF1 puncta (16.55 % ± 1.2%, $p < 0.0001$). This pattern was maintained following the one-hour recovery period (1Hr HS) (18.58% ± 1.3%, $P < 0.0001$ vs NHS). A similar pattern was observed for the small HSF1 foci. In the absence of heat shock (NHS), very few cells displayed small HSF1 puncta (1.49 % ± 0.37 %). Immediately following a heat shock (HSNR) and after the recovery period (1Hr HS), half of the cells exhibited small puncta (52.27 % ± 2.18 % and 50.19 % ± 1.68% respectively, both $p < 0.0001$ vs NHS). Together these results show that in the absence of heat shock, HSF1 rarely forms aggregates and that a one-hour heat shock results in the formation of large nuclear HSF foci in 15-20% of cells observed at that time point and small nuclear HSF foci in 50% of cells that are maintained for at least one hour following the return to control temperature.

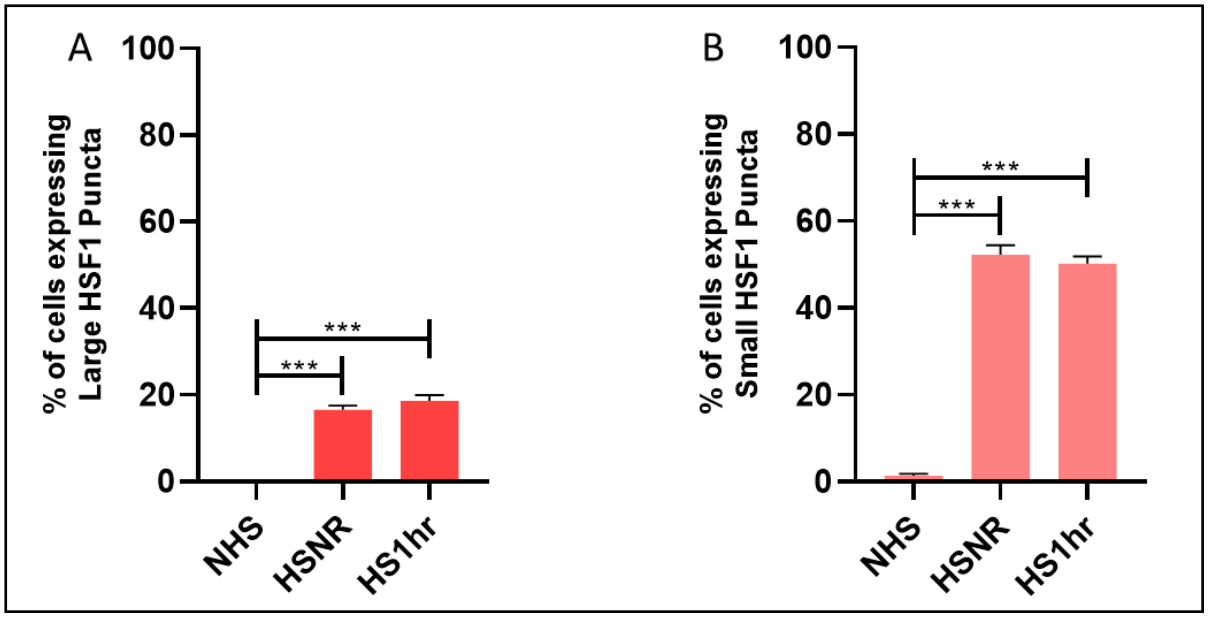

Figure 3. 13.Graphs showing the percentages of SH-SHSY cells expressing (A) large and (B) small HSF1 nSBs according to the MIA analysis. Cells were fixed and stained either without treatment (NHS), immediately after 1 hour of heat shock at 42° C (HSNR) or after 1 hour of recovery at 37° C following heat shock (1Hr HS). N=3 independent experiments. Each experiment is comprised by at least 6 wells per condition. A minimum of 10 images were included in the analysis per well. Lines represent mean +/- SEM. *=p<0.05, **=p<0.01, ***=p< 0.001.

We also analysed changes in SAFB protein distribution following heat stress and recovery using MIA. In the absence of HS (NHS), only a small portion of the cells exhibited large SAFB1 puncta (8.67%±0.73%) (Figure 3. 14 A). Immediately after heat shock (HSNR), there was a significant increase in the proportion of cells expressing large SAFB1 puncta (15.40% ± 1.98%, p=0.1373). However, after the recovery period (1Hr HS), SAFB1 distribution changed significantly compared to the control group (15.30% ± 1.8%, P=0.0049 vs NHS). SAFB1 did not co-localise with HSF1 in the absence of stress (0.01%±0.01%, N=6). However, a small but significant increase in SAFB1 co-localisation with HSF1 was seen immediately after an HS (4.87% ± 0.37%, p<0.0001) and following an HS plus 1 hr recovery (5.76%±1.72%, p=0.0441 (Figure 3. 14 B). Interestingly, the changes in SAFB1 distribution and co-localisation with HSF1 detected by the MIA analysis could not be observed by microscopic observation.

A different pattern was observed for the small SAFB1 puncta (Figure 3. 14 C). A considerable portion of the cell population expressed small SAFB1 puncta even under normal conditions (40.12%±1.71%). No significant changes were observed either immediately following heat shock (HSNR) or after the recovery period (1Hr HS) (52.32% ± 4.84% and 48.18% ± 3.75% respectively, both p>0.05 vs NHS). Despite the high percentage of cells expressing small SAFB1 puncta, no co-localisation between SAFB1 puncta and HSF1 nSBs could be found at any of the time points (Figure 3. 14 D).

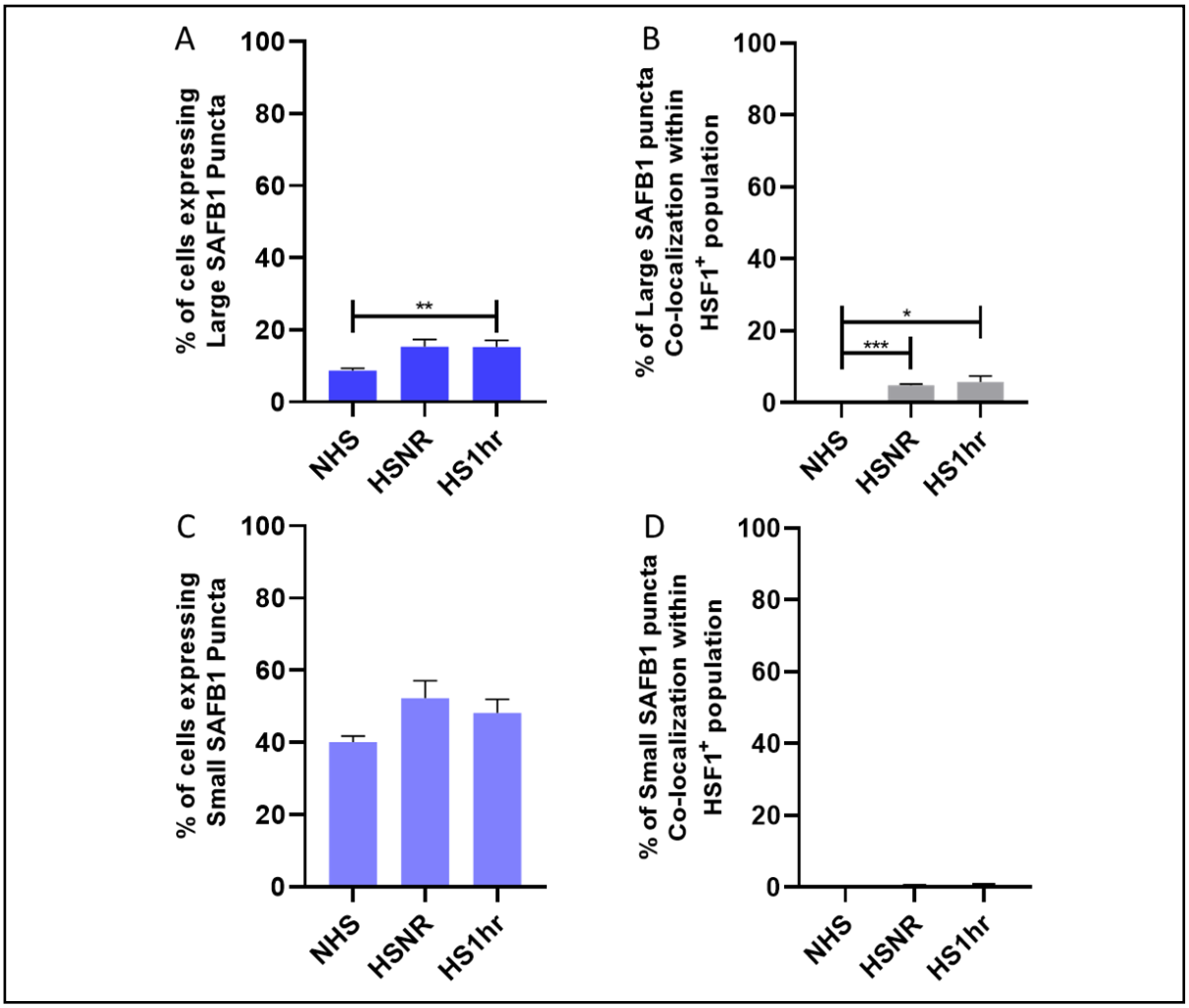

Figure 3. 14 Percentage of cells displaying Large and Small SAFB1 puncta and its co-localization within the HSF1 positive population. (A and C) Graphs showing the percentages of SH-SHSY cells expressing large (A) or small (C) SAFB1 puncta's; (B and D) graphs show the percentage of cells expressing (B) large or (D) small SAFB1 puncta in co-localization within HSF1 nSBs positive population, according to the MIA analysis. Cells were fixed and stained either without treatment (NHS), immediately after 1 hour of heat shock at 42° C (HSNR) or after 1 hour of recovery at 37° C following heat shock (1Hr HS). N=3 independent experiments. Each experiment is comprised by at least 2 wells per condition. A minimum of 10 images were included in the analysis per well. Lines represent mean +/- SEM. *=p<0.05, **=p<0.01, ***=p< 0.001.

The same parameters used to analyse SAFB1 were used in the SAFB2 experiments. In the no HS condition cells exhibited large SAFB2 puncta (21.96% ± 1.96%, N=6) (Figure 3. 15 A). Immediately after HS (HSNR), no significant changes in the proportion of cells expressing large SAFB2 puncta could be observed (18.75% ± 3.9%, N=9, $p>0.999$). However, after the 1 hr recovery period, a dramatic increase in the percentage of cells exhibiting SAFB2 puncta is seen (51.67% ± 2.21%, N=6, $p<0.0001$) in comparison to the control group. There is no co-localisation between SAFB2 and HSF1 prior to HS (0.02%±0.01%, N=6), and this increases to 62.06% ± 4.64% (N=9, $p<0.0001$) immediately after heat shock and 85.18%±2.45% (N=6, $p<0.0001$) following HS plus 1hour recovery (Figure 3. 15B).

The results show a large portion of the cell population express small SAFB2 puncta under normal conditions (81.27% ± 2.47%, N=6), which is significantly decreased immediately after heat stress (57.21%±7.59%, N=8, $p=0.0003$). Following the one-hour HS recovery period, this proportion increases significantly in comparison to the HSNR group (72.9% ± 7.77%, N=8, $p=0.026$ vs HSNR), returning to similar levels to the control group ($p=0.1844$ vs NHS) (Figure 3. 15C). Only a small amount of these SAFB2 puncta co-localise with HSF1 nSBs either immediately after heat stress (6.26%±0.99%, N=8, $p=0.0007$) or following recovery (7.9%±0.69%, N=8, $p<0.0001$). No small SAFB2/HSF1 co-localisation was observed in the absence of stress (Figure 3. 15D).

Taken together, these results show SAFB2 is recruited to the canonical large HSF1 nSBs following heat stress, whereas only a small percentage of SAFB1 is seen with HSF1 nSBs after HS.

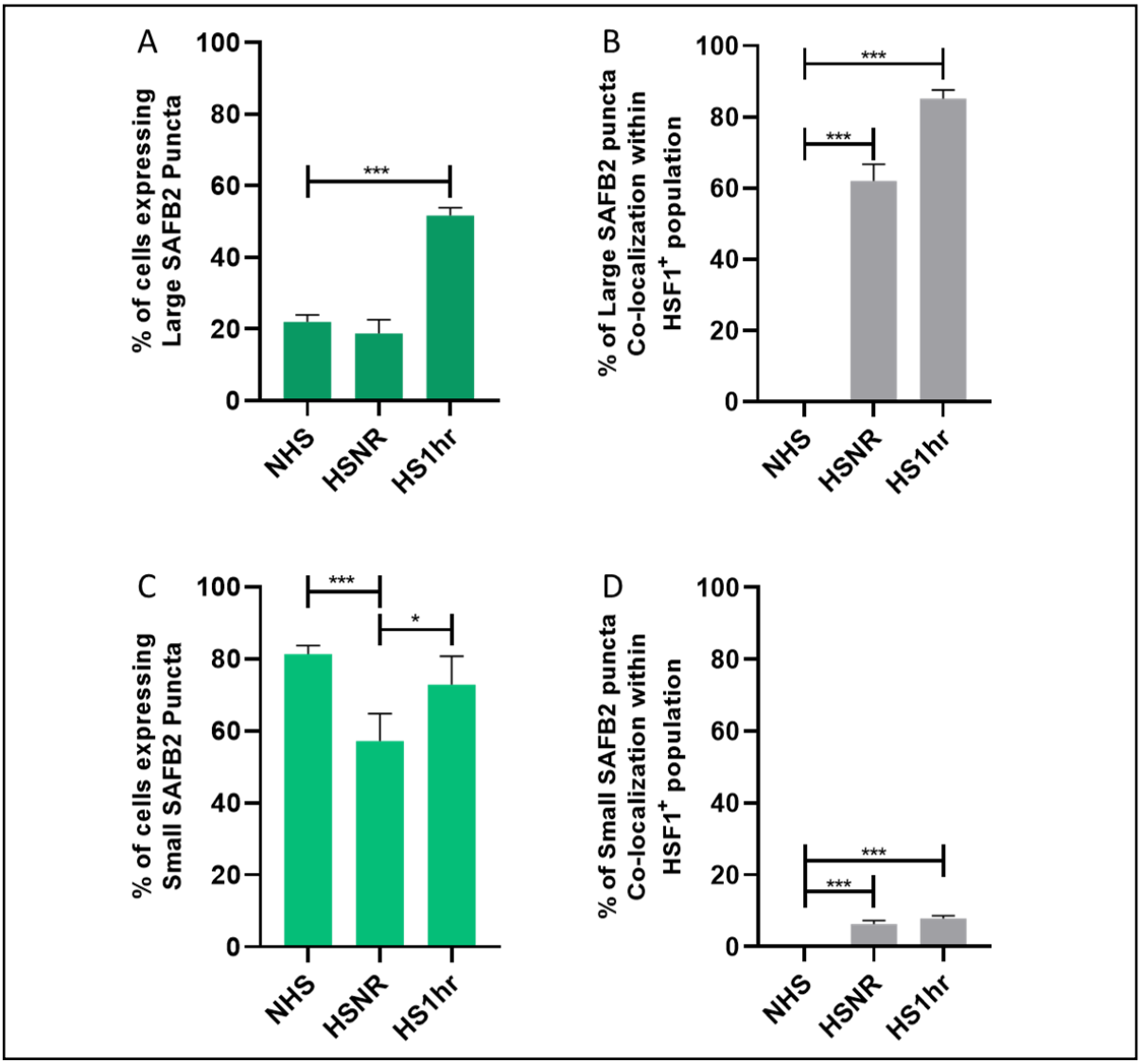

Figure 3. 15 Percentage of cells displaying large and small SAFB2 puncta and co-localization with HSF1. (A and C) Graphs showing the percentage of SH-SHSY cells expressing large (A) or small (C) SAFB2 puncta; (B and D) graphs show the percentage of cells expressing (B) large or (D) small SAFB2 puncta in co-localization within HSF1 nSBs positive population, according to the MIA analysis. Cells were fixed and stained either without treatment (NHS), immediately after 1 hour of heat shock at 42° C (HSNR) or after 1 hour of recovery at 37° C following heat shock (1Hr HS). N=3 independent experiments. Each experiment is comprised by at least 2 wells per condition. A minimum of 10 images were included in the analysis per well. Lines represent mean +/- SEM. *=p<0.05, **=p<0.01, ***=p< 0.001.

SLTM distribution following heat stress and recovery was also examined. No change was observed in the number of large SLTM puncta between the conditions examined (N=3, p>0.05) (Figure 3. 16. A). There was no co-localisation between large SLTM and HSF1 nSBs before and immediately after HS (0.06% ± 0.04%, N=4 and 1.5% ± 0.82%, N=5, p=0.2559 vs NHS, respectively), whereas there was an increase in the proportion of cells exhibiting co-localisation after the recovery period (22.85% ± 1.85%, N=3, p=0.0169 vs NHS) (Figure 3. 16 B).

Similarly, no changes were seen in the distribution of small SLTM puncta; the percentage of cells expressing small SLTM puncta (Figure 3. 16 C), and small SLTM/HSF1 co-localisation (Figure 3. 16 D) was unchanged across all time points. Less than 5% of the cell population exhibited small SLTM puncta (4.77% ± 1.95%, N=5) prior to heat shock, and this number was not significantly reduced in HSNR - 3.95% ± 1.06%, N=5, p>0.999 vs NHS, and 1Hr HS - 1.89% ± 0.72%, N=4, p=0.4545 vs NHS conditions. There was no co-localisation between small SLTM puncta and HSF1 foci (0.0%±0.00%, N=5 at NHS, HSNR or 1Hr HS, and p>0.999 for HSNR and 1Hr HS vs NHS).

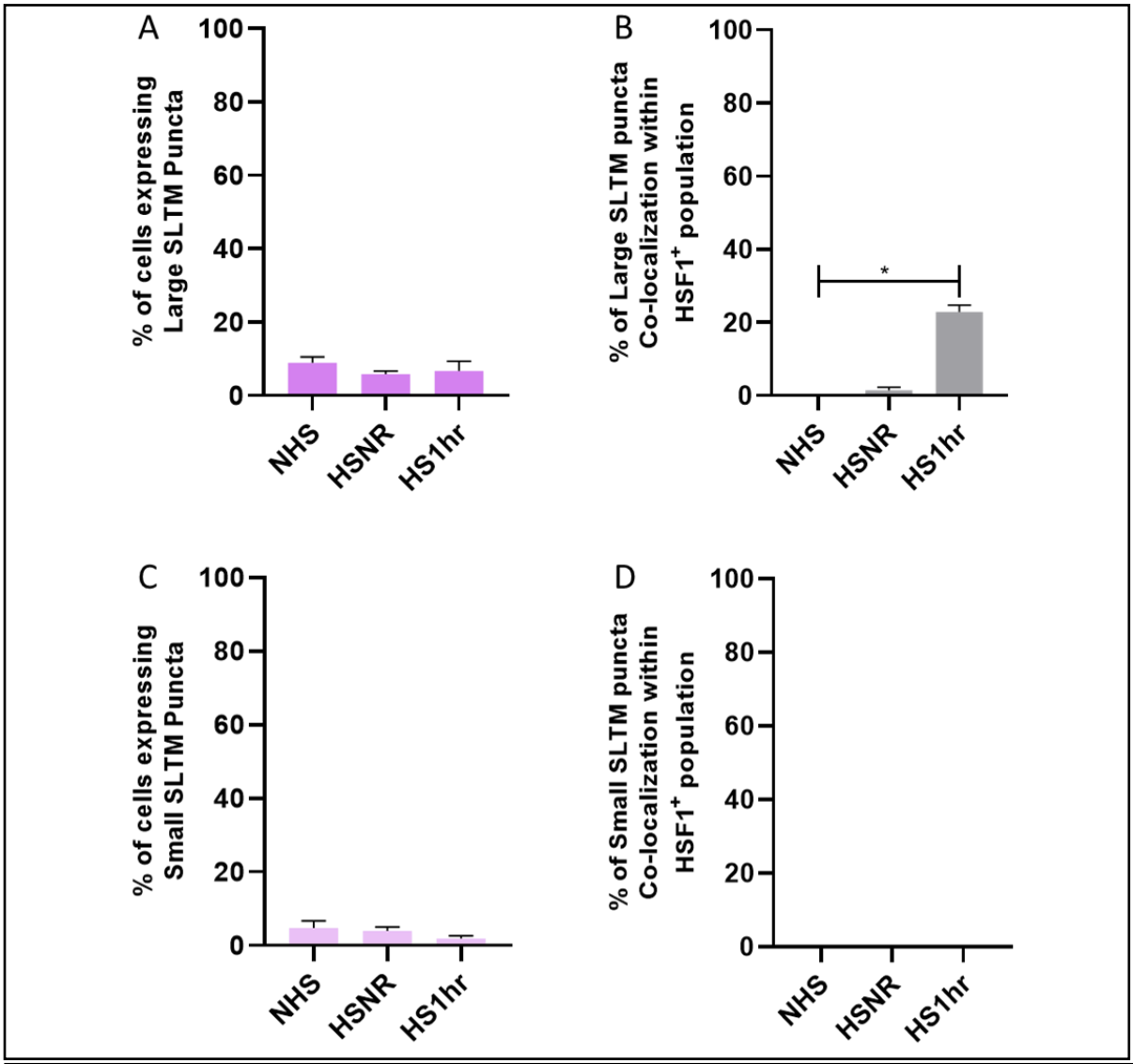

Figure 3. 16 Percentage of cells displaying Large and Small SLTM puncta and its co-localization within the HSF1 positive population. (A and C) Graphs showing the percentages of SH-SHSY cells expressing large (A) or small (C) SLTM puncta; (B and D) graphs show the percentage of cells expressing (B) large or (D) small SLTM puncta in co-localization within HSF1 nSBs positive population, according to the MIA analysis. Cells were fixed and stained either without treatment (NHS), immediately after 1 hour of heat shock at 42° C (HSNR) or after 1 hour of recovery at 37° C following heat shock (1Hr HS). N=3 independent experiments. A minimum of 10 images were included in the analysis per well. Lines represent mean +/- SEM. *=p<0.05, **=p<0.01, ***=p< 0.001.

3.2.7 RNA-Fluorescent *In Situ* Hybridisation

I next aimed to evaluate the recruitment of specific SAFB proteins to the sites of lncSATIII transcript accumulation. lncSATIII transcripts co-localise with HSF1^{+} nSBs [179, 200], and have been recently found bind to SAFB2 and SLTM via RRM domain [20]. Therefore, we expected to observe a similar recruitment pattern to that seen with HSF1 and described in the previous section. To characterise SAFB protein incorporation into the lncSATIII transcripts, *in situ* hybridisation was performed using biotinylated RNA probes against SATIII and BR322 (described in methods, section 2.10) to visualise lncSATIII foci in cells co-stained with SAFB1 or SAFB2. MIA analyses showed that HSF1 co-localised with large lncSATIII in approximately 40% of the lncSATIII positive cells, whereas SAFB2 co-localises with 50 to 60% of the large lncSATIII positive population following recovery (Figure 3. 17 to Figure 3. 20). However, no SAFB1 co-localised staining could be observed.

In the absence of heat shock (NHS), very few cells had large lncSATIII puncta (9.37%±1.04%). Immediately after heat shock (HSNR), there was a significant increase in the proportion of cells expressing large lncSATIII puncta (28.14%±2.36, $p=0.0011$), and this was maintained for 1 hour following the return to control temperatures (1Hr HS) (35.8%±5.27, $p<0.0001$) (Figure 3. 17). A similar trend was observed for the small lncSATIII foci. In the absence of HS (NHS), 21.93%±3.20% of the cells displayed small HSF1 puncta. Immediately following HSNR and after the recovery period (1Hr HS), a third of the cells exhibited small puncta (37.32%±2.26%, $p=0.0033$ and 39.75%±3.42%, $p=0.0008$, respectively). These results show that in the absence of HS SATIII, small puncta are more common in the non-heat stress condition, suggesting some background transcription. However, large foci of lncSATIII accumulation are only seen after HS, which clearly elicits a significant increase in the transcription of SATIII DNA.

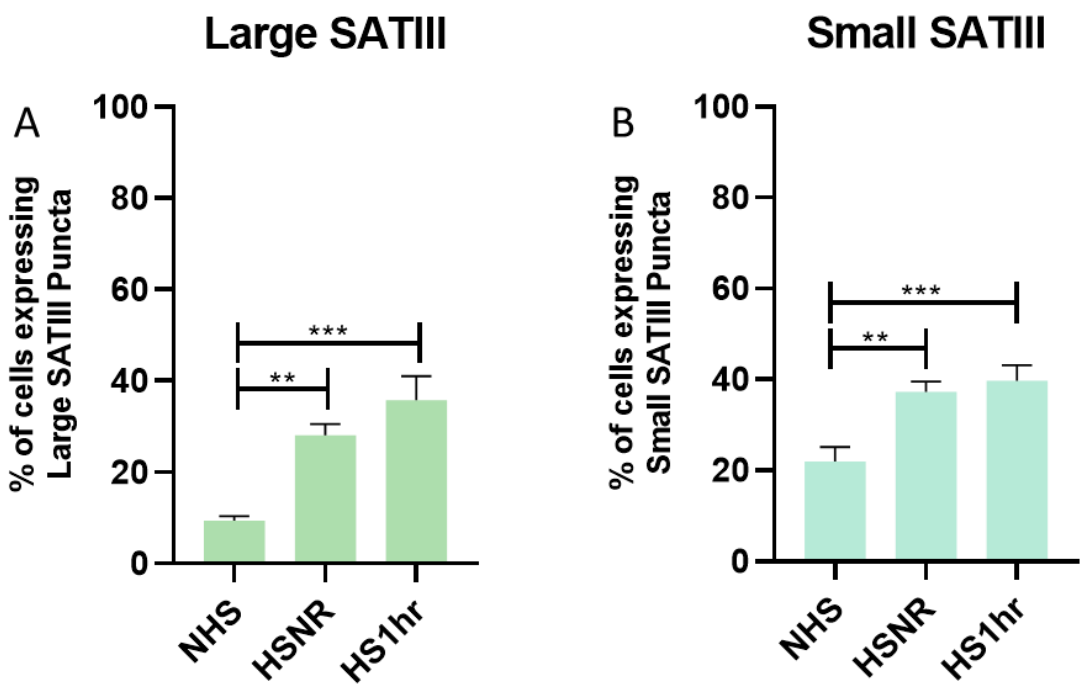

Figure 3. 17 Percentage of SH-SHSY cells displaying (A) large and (B) small lncSATIII puncta (please be consistent when using lower and upper case, you switch between the two and the external will make you correct these) , according to the MIA analysis. Cells were fixed and stained either without treatment (NHS), immediately after 1 hour of heat shock at 42° C (HSNR) or after 1 hour of recovery at 37° C following heat shock (1Hr HS). N=3 independent experiments. A minimum of 10 images were included in the analysis per well. Lines represent mean +/- SEM. *=p<0.05, **=p<0.01, ***=p< 0.001

In the absence of stress, the lncSATIII presented a diffuse dotted pattern, across the nucleus and cytoplasm. However, the SATIII staining is more pronounced within the cytoplasm (Figure 3. 18 A – top row), which is suggestive of non-specific staining. After stress induction, synthesis of lncSATIII can be observed through the formation of 2 large foci easily distinguished within the nuclear compartment (Figure 3. 18A – middle row), which persists in site after 1 hour of recovery under normal conditions (Figure 3. 18 A – bottom row). The percentage of cells displaying co-localisation between large lncSATIII puncta and HSF1 increases by approximately 30% (1.21%±0.52% - 27.64%±3.2%, $p<0.0001$) immediately after heat shock and by 40% (39.23%±1.82% - $p<0.0001$ vs NHS) in the one-hour recovery period. The co-localisation of small lncSATIII and HSF1(Figure 3. 18) is also increased from a zero baseline after stress induction (7.37%±0.58%, $p<0.0001$) and following the recovery period (10.42%±1.16%, $p<0.0001$).

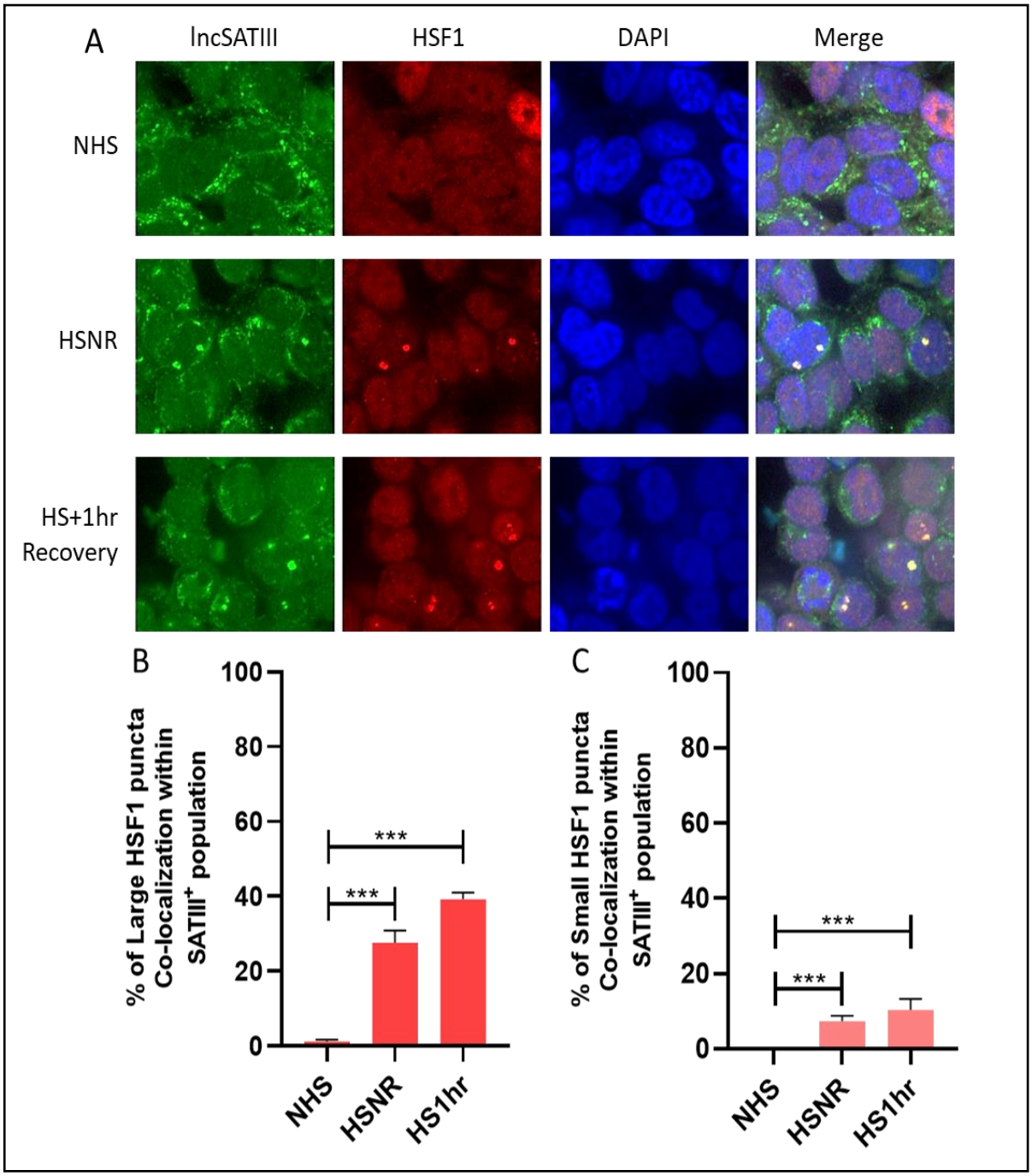

Figure 3. 18 Representative images of lncSATIII RNA-FISH and HSF1 ICC staining and subsequent Image analysis. (A) SH-SY5Y Cells were submitted to no stress, one hour of heat stress or one hour of heat stress followed by one hour of recovery prior to fixing. Staining for lncSATIII RNA followed by HSF1 ICC was carried out with specific antibodies. Hoechst was used for nuclear identification. Merge images of lncSATIII and HSF1 shows good co-localization between the nSBs. Water 60x objective, 5x Zoom. (B and C) Percentage of SH-SHSY cells displaying (B) Large and (C) Small lncSATIII-HSF1 co-localization, according to the MIA analysis. N=3 independent experiments. A minimum of 10 images were included in the analysis per well. Lines represent mean +/- SEM. *=p<0.05, **=p<0.01, ***=p< 0.001.

Experiments to investigate the localisation of SAFB1 with nSBs were also carried out. However, the conditions needed for co-staining altered resulted in a granular pattern of SAFB1 distributed across both nuclear and cytoplasmic (Figure 3. 19). This was somewhat different to that observed previously (Figure 3. 9 and Figure 3. 10) as although nuclear staining was present, it was less prominent. There was also no apparent co-localisation between the SATIII transcripts and SAFB1 after heat stress induction. Overall, the results also show no significant change in the percentage of cells displaying co-localisation between large lncSATIII puncta and SAFB1 (from 18.54%±0.53% to 12.38%±0.96%, p=0.0096 and 13.46%±0.82%, at HSNR and 1Hr HS respectively) (Figure 3. 19). These results could possibly be explained by using formamide reagent, present in the hybridisation buffer for the RNA-FISH technique (see methods). The RNA staining is carried out before the ICC to avoid RNA degradation. Formamide is a desaturating agent that increases the instability of nucleic acids, expanding probe access to DNA/RNA and facilitates the hybridisation process at lower temperatures [220]. Therefore, it is an imperative reagent. However, as SAFB1 directly binds with RNA and proteins that interact with RNA, the formamide reagent will likely reduce these interactions and thus alter SAFB1 staining.

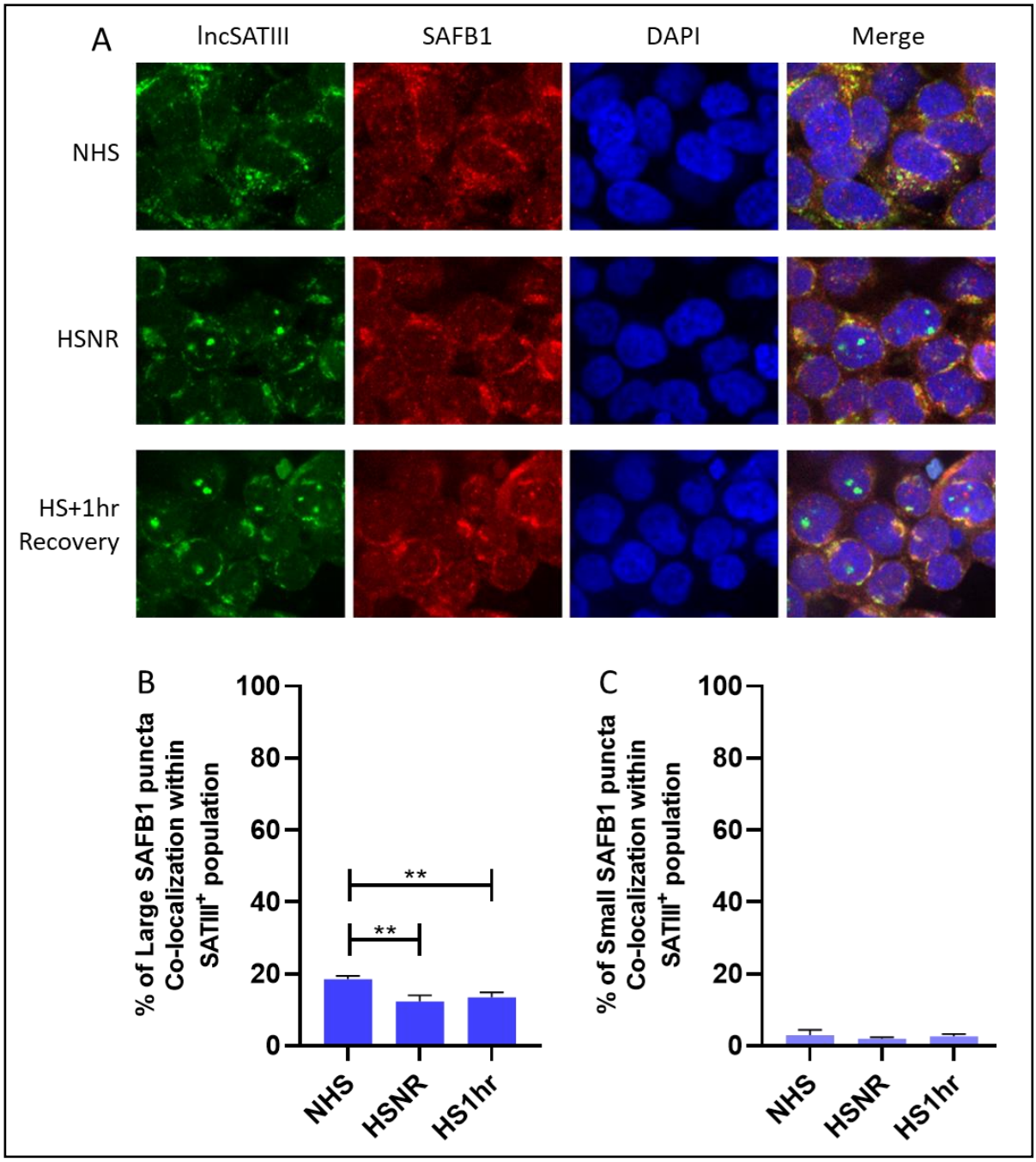

Figure 3. 19 Representative images of lncSATIII RNA-FISH and SAFB1 ICC staining and Image analysis results. (A) SH-SY5Y Cells were submitted to no stress, one hour of heat stress or one hour of heat stress followed by one hour of recovery prior to fixing. Staining for lncSATIII RNA followed by SAFB1 ICC was carried out with specific antibodies. Hoechst was used for nuclear identification. Merge images of lncSATIII and SAFB1 shows no co-localization between the proteins. Water 60x objective, 5x Zoom. (B and C) Percentage of SH-SHSY cells displaying (B) Large and (C) Small lncSATIII-SAFB1 co-localization, according to the MIA analysis. N=3 independent experiments. A minimum of 10 images were included in the analysis per well. Lines represent mean +/- SEM. *=p<0.05, **=p<0.01, ***=p< 0.001.

Following SAFB1 analysis, SAFB2 distribution and SATIII co-localisation was investigated. The percentage of cells displaying co-localisation between large lncSATIII puncta and SAFB2 (Figure 3. 20 B) remained unaltered following an HS (32.77%±2.17% - 33.40%±8.9%, p=0.9975). Following the one-hour recovery period, however, the percentage of cells showing co-localisation between lncSATIII and SAFB2 increased (55.48%±6.78, p=0.1122) though this was not statistically significant (Figure 3. 20 C). In addition, there was a significant decrease in the presence of small SAFB2 puncta following an HS. The images shown in Figure 3. 20 Representative images of lncSATIII RNA-FISH and SAFB2 ICC staining and Image analysis results. (A) SH-SY5Y Cells were submitted to no stress, one hour of heat stress or one hour of heat stress followed by one hour of recovery prior to fixing. Staining for lncSATIII RNA followed by SAFB2 ICC was carried out with specific antibodies. Hoechst was used for nuclear identification. Merge images of lncSATIII and SAFB2 shows higher co-localization between the protein and lncSATIII. Water 60x objective, 5x Zoom. (B and C) Percentage of SH-SHSY cells displaying (B) Large and (C) Small lncSATIII-SAFB2 co-localization, according to the MIA analysis. N=3 independent experiments. A minimum of 10 images were included in the analysis per well. Lines represent mean +/- SEM. *=p<0.05, **=p<0.01, ***=p< 0.001. are representative of the results obtained and clearly show SAFB2 is co-locating with SATIII puncta. However, as large SAFB2 puncta exist in control of HS conditions, the algorithm cannot detect a difference between the conditions.

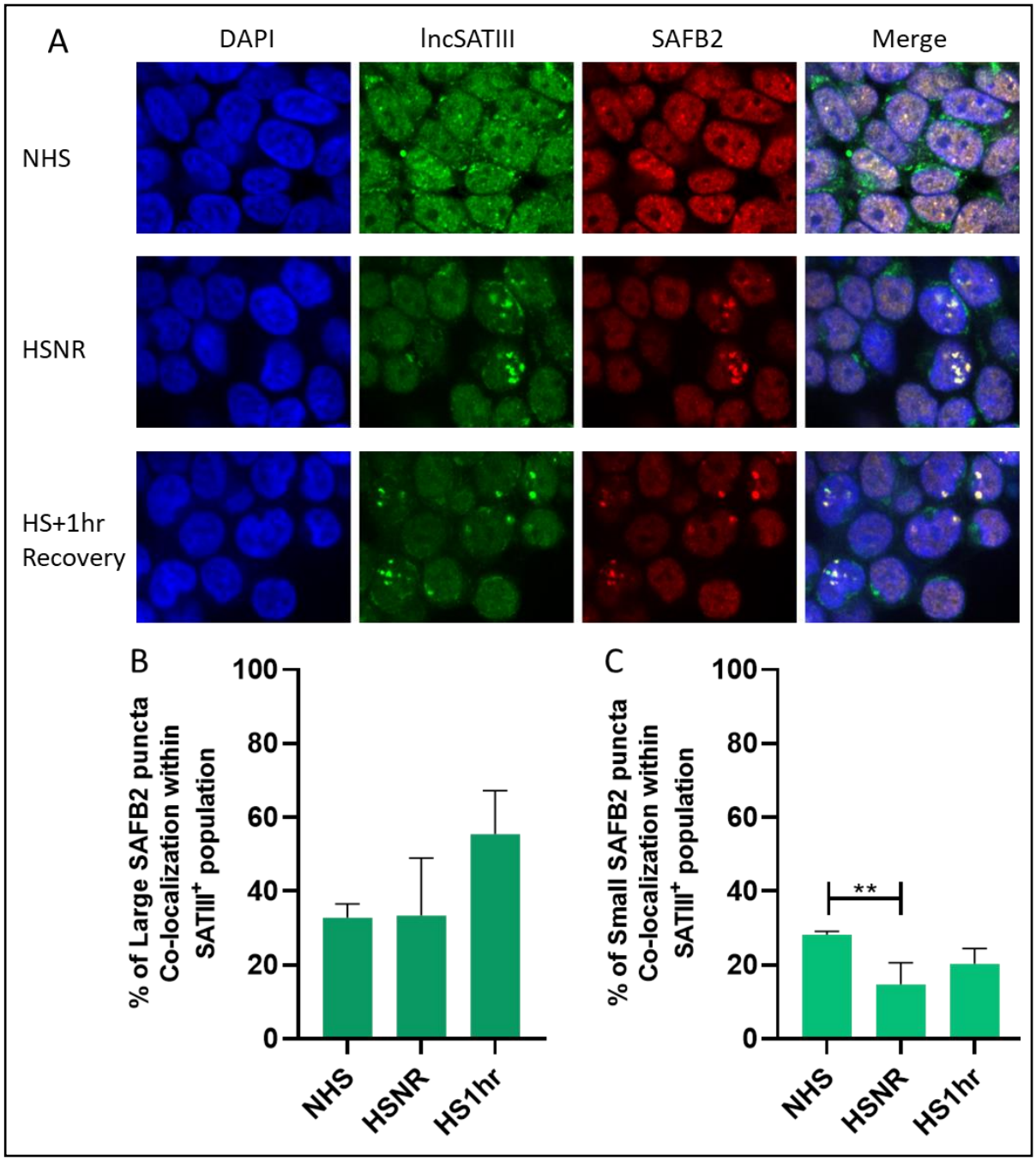

Figure 3. 20 Representative images of lncSATIII RNA-FISH and SAFB2 ICC staining and Image analysis results. (A) SH-SY5Y Cells were submitted to no stress, one hour of heat stress or one hour of heat stress followed by one hour of recovery prior to fixing. Staining for lncSATIII RNA followed by SAFB2 ICC was carried out with specific antibodies. Hoechst was used for nuclear identification. Merge images of lncSATIII and SAFB2 shows higher co-localization between the protein and lncSATIII. Water 60x objective, 5x Zoom. (B and C) Percentage of SH-SHSY cells displaying (B) Large and (C) Small lncSATIII-SAFB2 co-localization, according to the MIA analysis. N=3 independent experiments. A minimum of 10 images were included in the analysis per well. Lines represent mean +/- SEM. *=p<0.05, **=p<0.01, ***=p< 0.001.

Control images were acquired using an oligo-control probe complementary to the non-targeting control pBR322 sequence and a sense SATIII probe. Samples were submitted or not to one hour of heat stress followed by one hour of recovery at 37° C and processed following the RNA-FISH protocol described in methods.

Figure 3. 22 Figure 3. 22 and Figure 3. 21 show that the SATIII probe and the pBR322 probe can hybridise within the cells' cytoplasm, producing a non-specific background staining. However, no intra-nuclear staining could be observed in any images under non-HS or HS with either probe. For the purpose of these experiments, only nuclear SATIII staining was counted, and this was only detected by the anti-sense SATIII probe, and hence the RNA-FISH assay shows a high degree of specificity.

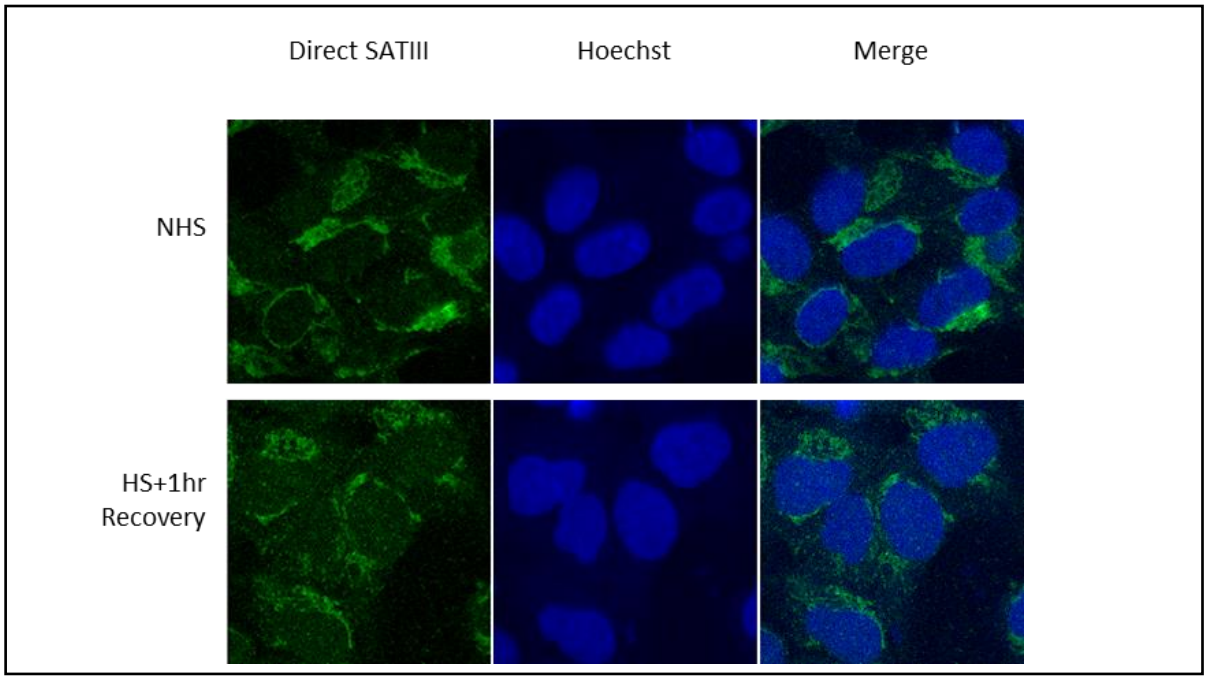

Figure 3. 22 Images of SH-SY5Y cells stained with a control sense probe against SATIII. Cells were submitted to either No Heat Shock or 1hr of Heat Shock followed by 1hr of Recovery prior to the fixing. Oil 63x objective.

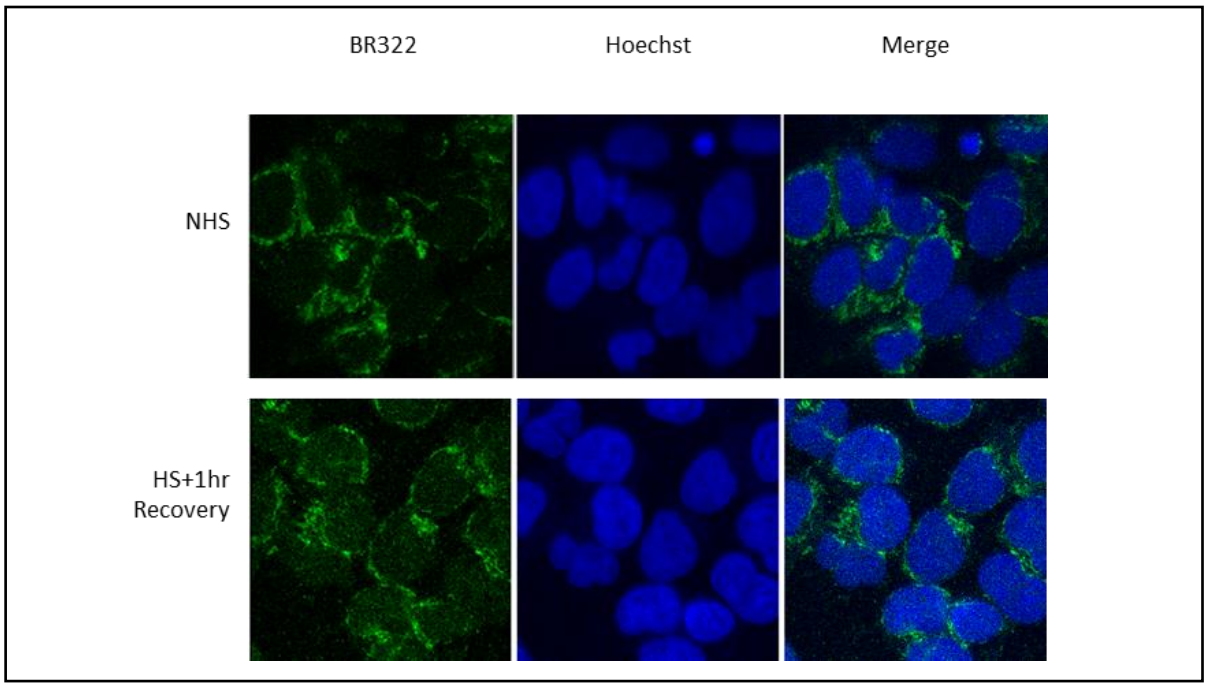

Figure 3. 21 Images of SH-SY5Y cells stained for the non-targeting control BR3222. Cells were submitted to either No Heat Shock or 1hr of Heat Shock followed by 1hr of Recovery prior to the fixing. Oil 63x objective.

These results suggest that SAFB2 is strongly associated with lncSATIII transcripts, and this was not inhibited by treatment with formamide. This suggests a stronger association between SAFB2 and the SATIII sequences or that SAFB2 is strongly associated with SATIII binding proteins (e.g. HSF1).

3.2.8 SATIII Quantitative RT-PCR (qRT-PCR)

SAFB 1 is known to bind GAAGA motifs [30] with high affinity while SAFB2 and SLTM have been reported to bind repetitive sequences within SATIII transcripts. Considering SAFB proteins bind RNA polymerase we hypothesised that they may regulate the transcription of lncSATIII RNA. To investigate this hypothesis, we knocked down SAFB1 and/or SAFB2 using siRNAs and measured lncSATIII levels using rt-qPCR (Methods, sections 2.7, 2.13). In the absence of heat stress, no significant changes in lncSATIII levels were observed in any conditions (siSAFB1, siSAFB2 and siSAFB1SAFB2 treated, $p>0.99$ vs siNTC at NHS time point – Figure 3. 23 A). Immediately following one hour of HS (HSNR), there was an increase in the SATIII lncRNA levels that was not significant when compared to the NHS time point ($p>0.5$ vs NHS siRNA equally treated). Additionally, SAFB KDs did not significantly alter the expression levels of lncSATII within the HSNR time point ($p>0.8$ vs siNTC at HSNR). However, lncSATIII levels increased significantly after 1hr of recovery following heat stress (1Hr HS) for all siRNA treated samples in comparison to the NHS group (each comparison reached a significance of at least $p<0.05$ vs NHS siRNA - Figure 3. 23A). The knockdown of SAFB1 and double SAFB1/SAFB2 knockdown did not alter lncSATIII levels (896±180 and 591±68fold change vs siNTC, respectively, $p>0.2$). However, the single SAFB2 KD resulted in a sharp increase in lncSATIII levels at 1hr HS compared to the siNTC control (1725±460 fc, $p=0.0183$). lncSATIII levels in the SAFB2 depleted cells were also significantly increased in comparison to SAFB1 KD ($p=0.0131$) and SAFB1/SAFB2 KD ($p<0.001$) (Figure 3. 23A). These results suggest that SAFB2 may modulate SATIII transcription. Interestingly, the double SAFB1/SAFB2 knockdown did not display the same effect. Previous work has shown that the knockdown of SAFB1 elicits a compensatory increase in SAFB2 while knocking down SAFB2 may increase SAFB1 expression. Therefore, it is possible that a compensatory increase in SAFB1 levels alters lncSATIII transcription in the absence of

SAFB2 (mRNA and/or protein). However, although we see an increase in SAFB1 mRNA levels following SAFB2 KD (Figure 3. 23), this did not reach significance.

Heat stress treatments did not alter basal levels of SAFB1 mRNA or SAFB2 in siRNA treated control cells (SAFB1: p=0.504 NHS vs HSNR and p=0.282 NHS vs 1Hr HS; SAFB2: p=0.790 NHS vs HSNR and p=0.051 NHS vs 1hr HS) (Figure 3. 23 B and C). However, SAFB1 and SAFB2 mRNA levels were reduced significantly in all of the KDs treatment groups (SAFB1: no HS =0.483±0.116 fc, vs siNTC-NHS p=0.0014; HS=0.232±0.026 fc, p<0.0001 vs siNTC-HSNR; 1hr HS, R =0.514±0.131 fc, p<0.0001 vs siNTC-1Hr HS - Figure 3. 23 B. SAFB2: 0.482±0.081 fc, p=0.003 vs siNTC-NHS; 0.418±0.039 fc, p<0.0001 vs siNTC-HSNR; 0.472±0.070 fc, p<0.0001 vs siNTC-1Hr HS - Figure 3. 23C).

SAFB1 and SAFB2 were also individually KD in the double KD experiments reaching a significance of at least p=0.001 in all conditions. Additionally, there were no significant variations in SAFB1 or SAFB2 mRNA levels in the siRNA treated cells between conditions (Figure 3. 23 B).

Figure 3. 23 D shows the effects of SAFB1, SAFB2, SAFB1/SAFB2 depletion and control (siNTC) on SLTM mRNA levels. SLTM mRNA levels were measured to see if SAFB1/2 KD elicited a compensatory response. Heat stress treatments did alter basal levels of SLTM mRNA in siRNA control samples immediately after heat stress (0.781±0.063 fc, p=0.047 NHS vs HSNR), but not following one-hour recovery (0.833±0.081 fc, p=0.164 NHS vs 1Hr HS). SLTM mRNA levels remained unchanged upon SAFB1 and SAFB2 depletion at all time points. SLTM levels were observed to fall following an HS, and knockdown of SAFB1/2 did not alter this effect.

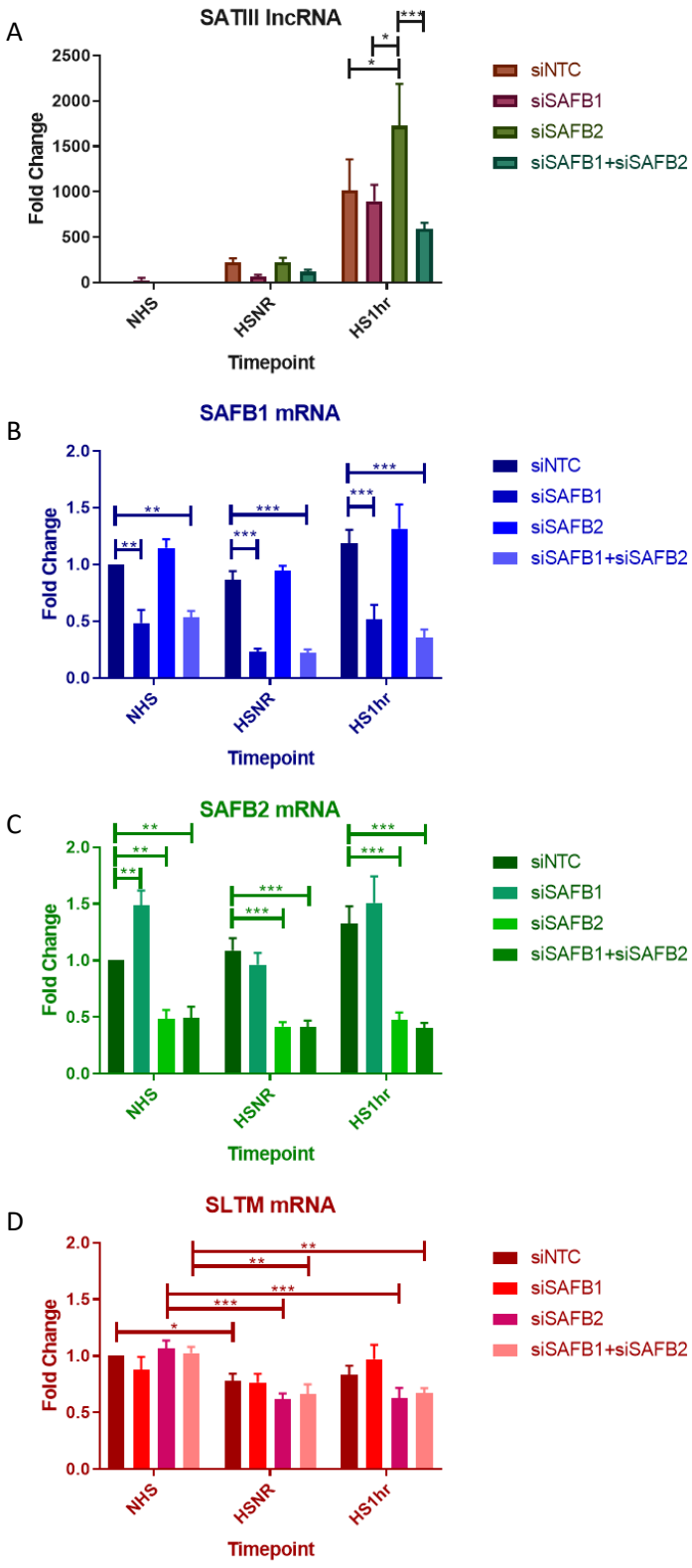

Figure 3. 23 mRNA expression levels of lncSATIII, SAFB1, SAFB2 and SLTM upon SAFB1 and/or SAFB2 depletion. SH-SY5Y cells were treated with 60 pM of siRNA against either SAFB1, SAFB2, SAFB1+SAFB2 or NTC and submitted to heat shock treatments (NHS, HSNR or HS+1hr recovery). mRNA was collected and measured by Quantitative RT-PCR. Measurements were made in 3 separately experiments. Values are means ± SEM, statistical analysis was performed by ANOVA and post-hoc t-tests *P <0.05, **P <0.01, ***P <0.001.

3.2.9 SAFB proteins distribution in primary fibroblasts

Jolly and collaborators (1997) reported that the number of cells expressing HSF1^{+} nSBs declined with increasing cell passage [200]. Our studies were conducted on cancer lines with a high passage number, and we therefore carried out HSF1 and SAFB1 immunostaining on heat shocked primary human fibroblasts that had only been passaged four to eight times. Fibroblasts were fixed following staining and three independent experiments were quantified by manual counting of the stress bodies. SLTM staining was only carried once and therefore no statistical analyses could be conducted.

We measured the percentage of fibroblasts with HSF1 nSBs, in the absence of stress, immediately after one hour of heat stress and after one hour of recovery following one hour of heat stress. Figure 3. 24 shows that in primary fibroblasts, the HSF1 nSBs are formed in 89.2±3.58% of the cells immediately after heat stress ($p<0.001$ vs NHS) and in 67.7±4.82% of the fibroblasts submitted to recovery following heat stress ($p<0.001$ vs NHS). Representative images of fibroblasts stained for HSF1 are shown in the subsequent Figure 3. 25 and Figure 3. 26.

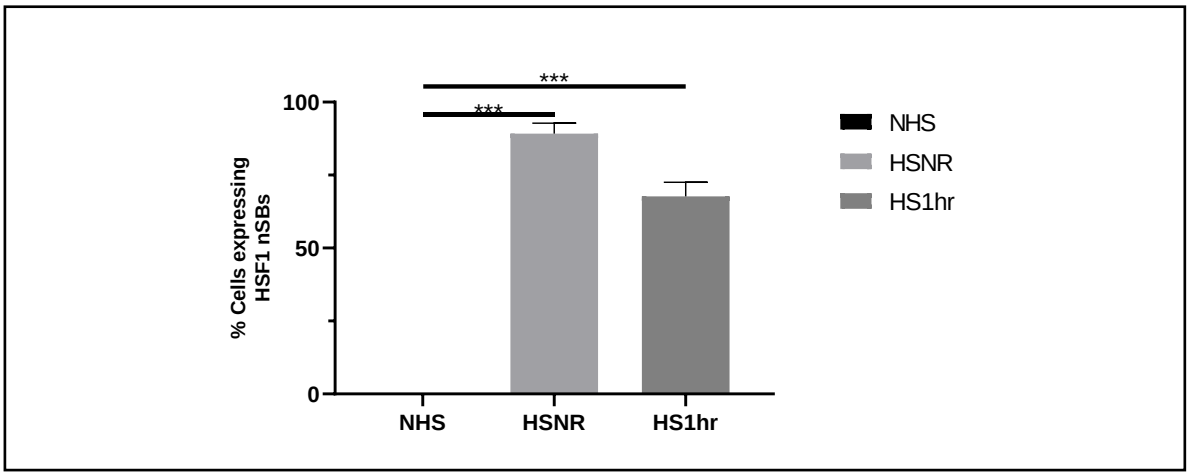

Figure 3. 24 Percentage of fibroblasts exhibiting HSF1 nSBs before and after heat stress, and recovery.

Figure 3. 25 shows fibroblasts stained for SAFB1 (green) and HSF1 (red) in panel A. SAFB1, shown in the literature as a nucleic protein, can be seen in smaller amounts in

the cytoplasm as well. SAFB1 displays a slight re-distribution within the nucleus following heat shock and recovery. However, SAFB1 is not recruited to the HSF1 nSBs, as shown in Figures 3. 28 B and C. No significance could be driven due to the lack of values observed.

On the other hand, SAFB2 presents a different distribution pattern following heat stress. Figure 3. 26A shows the SAFB2 (green) and HSF1 (red) distribution in fibroblasts following a stress response. SAFB2 formed puncta in 37.1±12.8 % of the fibroblasts immediately after heat shock ($p=0.02$ vs NHS). This percentage increased to 66.6±19.8% after recovery ($p=0.01$ vs NHS) (Figure 3. 26 B). SAFB2 puncta are also co-localised with the same HSF1 nSBs, showing that SAFB2 is recruited to the nSBs following heat stress. Immediately after heat stress, 39.4±11.8% of the HSF1+ fibroblast population ($p=0.01$ vs NHS) showed co-localisation between SAFB2 and HSF1, whereas after recovery following heat stress, 70.6±26.4% of the same population exhibited SAFB2-HSF1 co-localisation (Figure 3. 26 C).

Next, we stained the fibroblasts for SLTM and HSF1 (Figure 3. 27). SLTM is known to be present mostly in the cytoplasmatic compartment, as can be seen in Figure 3. 27. SLTM did not change its distribution in both time points following heat shock, neither co-localised with HSF1. SLTM staining was conducted in only one independent experiment with no repeats. Therefore, no statistics could be generated.

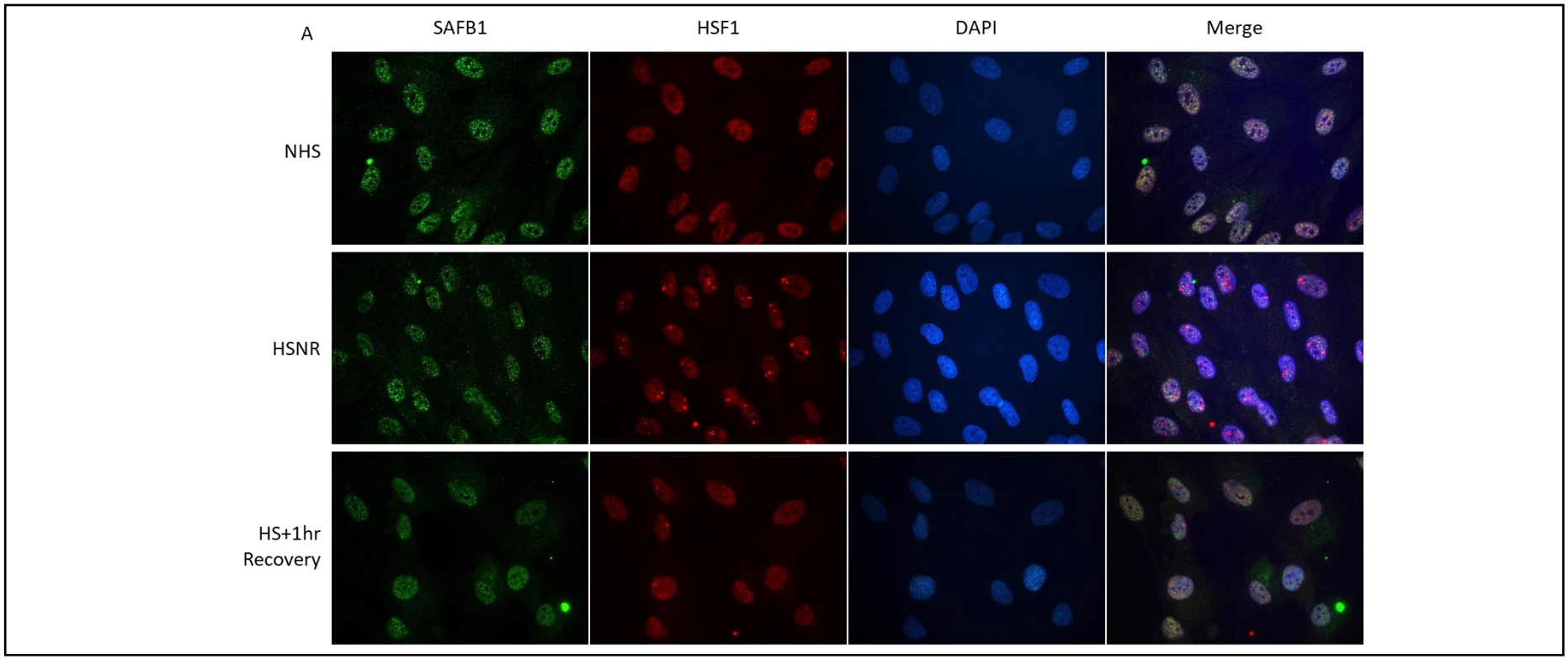

Figure 3. 25 Distribution of SAFB1 and HSF1 in Human primary fibroblasts

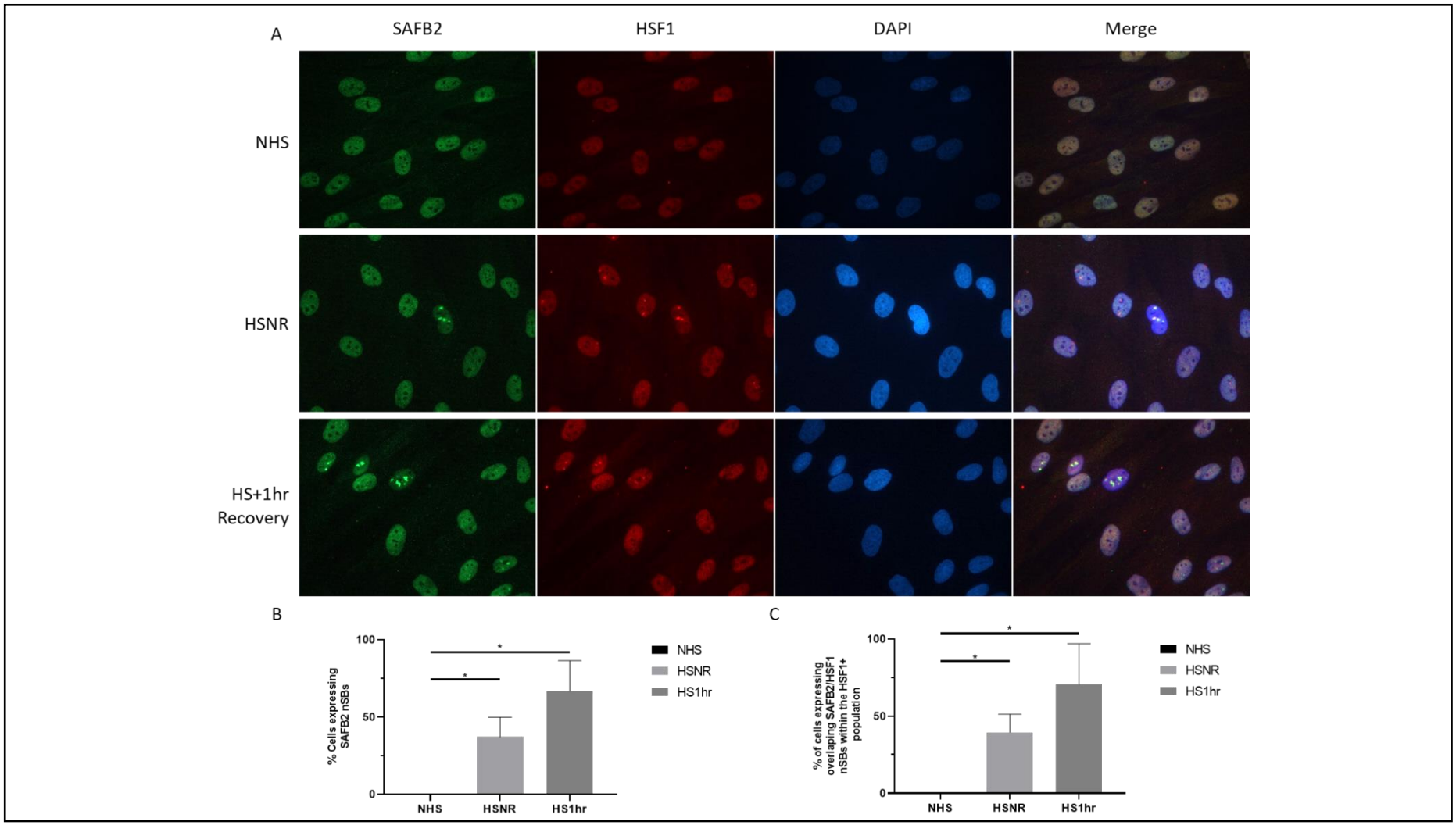

Figure 3. 26 distribution of SAFB2 and HSF1 in human fibroblasts

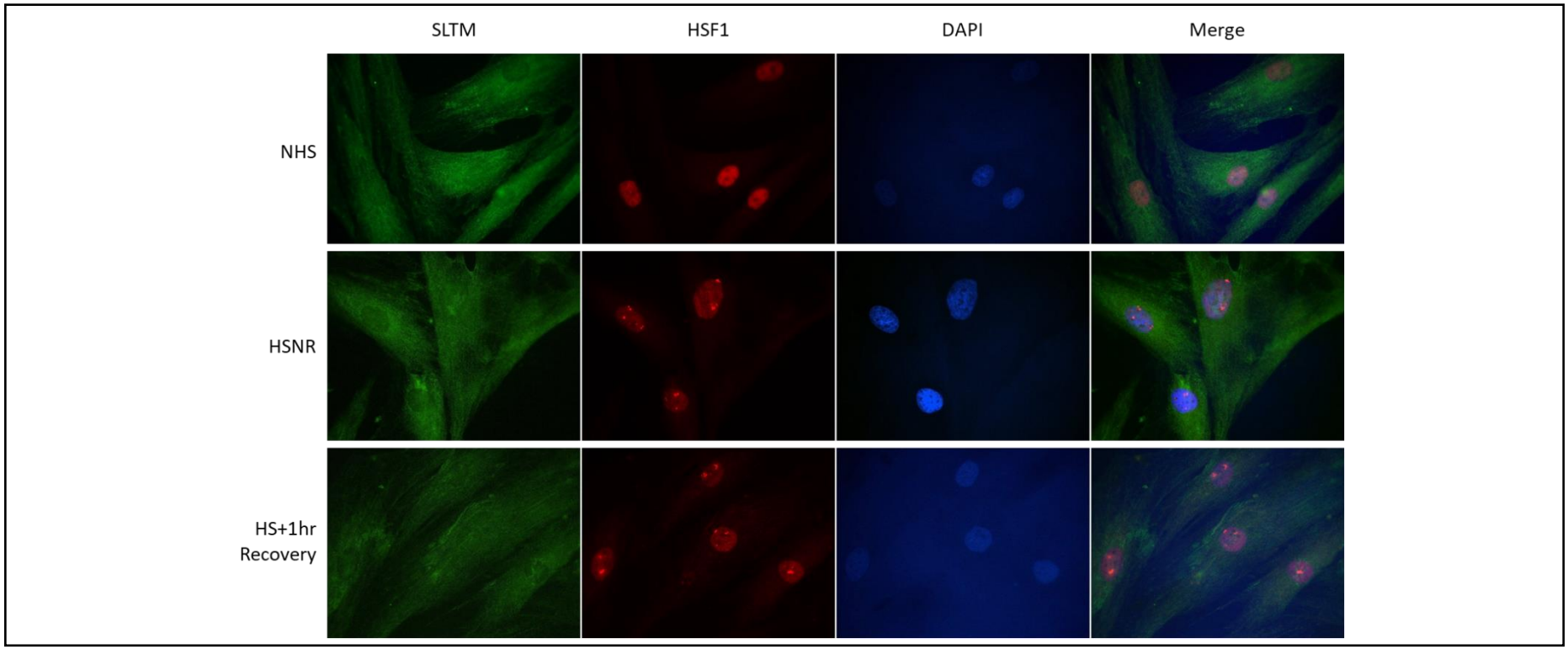

Figure 3. 27 SLTM distribution in human fibroblasts

3.2.10 Fibroblast SATIII Images

We also evaluated the expression of SATIII RNA in primary human fibroblasts. We used fluorescent in vitro hybridisation (FISH) with an RNA probe to stain the long non-coding SATIII RNA formed following heat stress in the human fibroblasts (methods section 2.2.3, 2.4, 2.9-2.10). Figure 3. 28 shows that no SATIII RNA is expressed in the nucleus in the absence of heat stress, whereas SAFB1 is evenly distributed across the compartment. The lncSATIII foci formed within one hour of heat stress and persisted for more than two hours after heat stress. SAFB1 remains unchanged up to one hour after heat stress. Interestingly, after longer recovery times (two hours of recovery following heat stress), SAFB1 started to form aggregates within the nucleus and these co-localised with lncSATIII nSBs.

In accordance with the previous results in section 3.2.4, in Figure 3. 29, it can be seen that SAFB2 starts to form aggregates within one hour of heat stress induction, and these aggregates persist for at least two hours after heat stress is terminated. The number of cells exhibiting SAFB2 puncta increases with time until almost all fibroblasts display SAFB2 puncta. However, this refers to microscopic observations since no measurements nor counting were performed. Interestingly, the observed co-localisation between SAFB2 and the lncSATIII is almost complete at all time points

In order to verify the robustness of the SATIII staining, images were generated using an RNA non-targeting control probe, the BR322, and a direct probe targeting SATIII sequences in the genome. Figure 3. 30 shows that both probes failed to generate a light signal that could be detected by microscopy before and after heat exposure, validating the lncSATIII staining observed in Figure 3. 28 and Figure 3. 29.

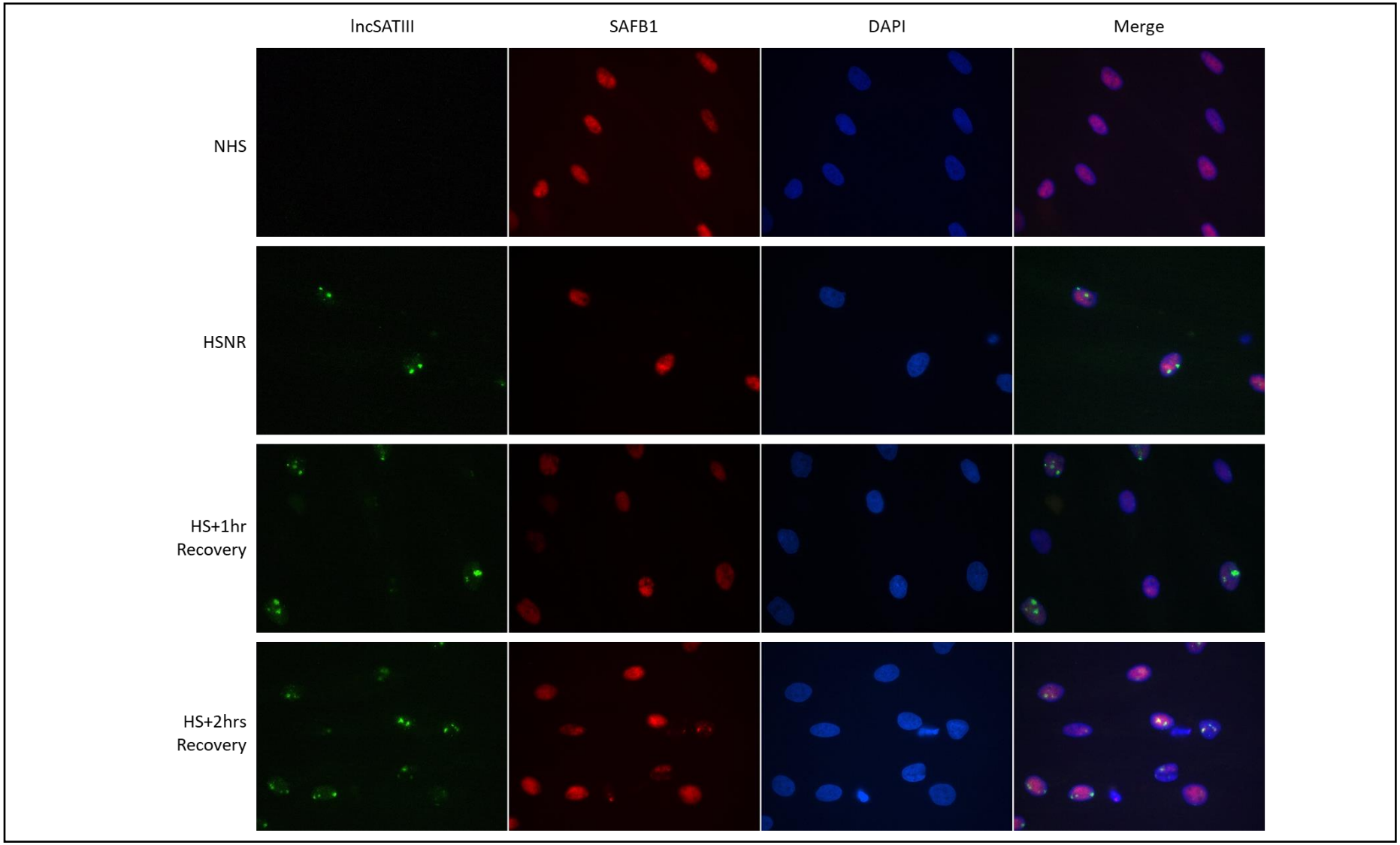

Figure 3. 28 lncSATIII and SAFB1distribution in human fibroblasts

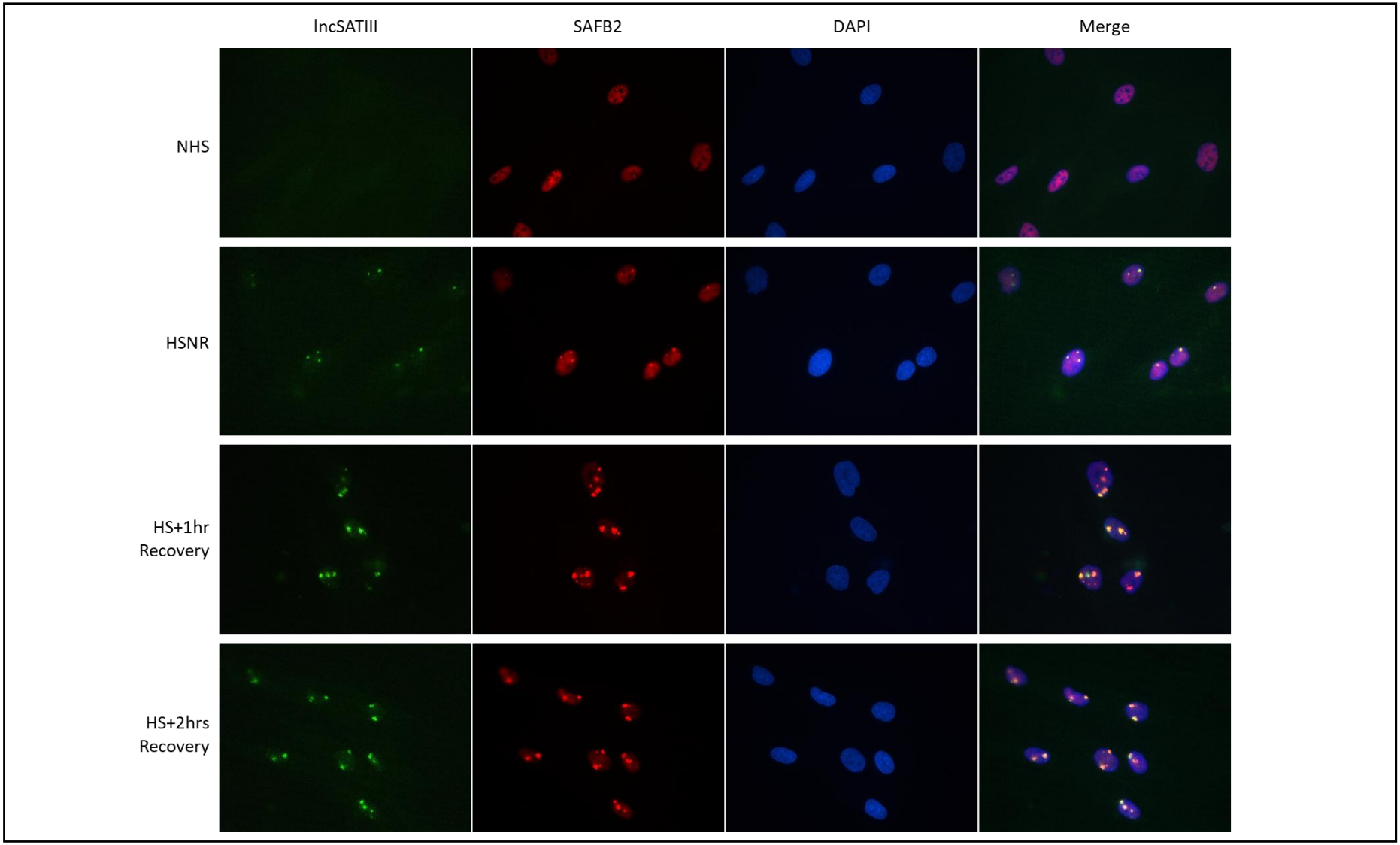

Figure 3. 29 lncSATIII and SAFB2 distribution in human fibroblasts

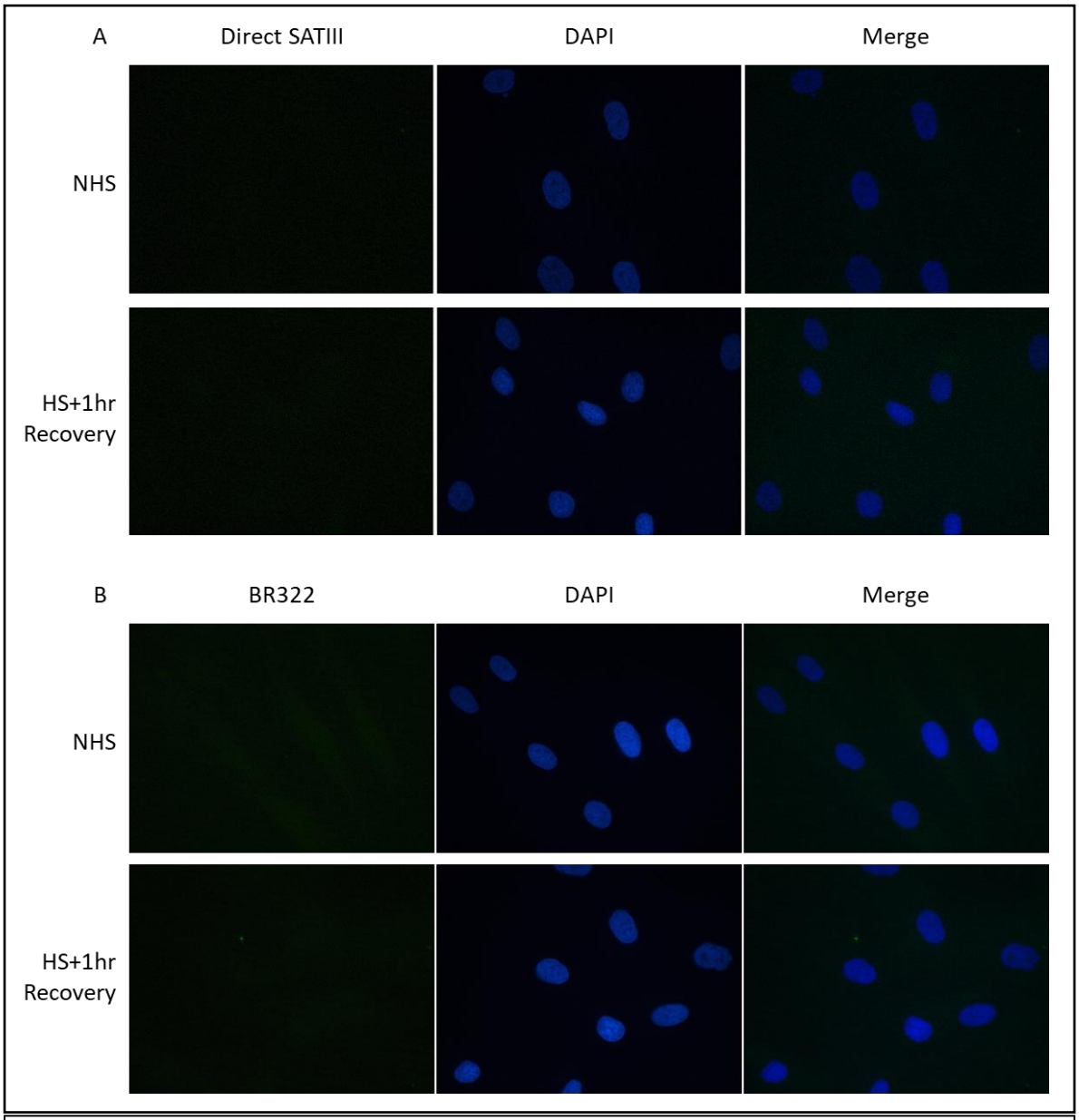

Figure 3. 30 Fibroblasts hybridized with RNA probes against genomic SATII repeats (A) and non-targeting BR322 (B), before and after heat stress and recovery time

3.3 Discussion

3.3.1 Key findings

We used antibodies specific for SAFB1 and SAFB2 to investigate their recruitment to nSBs following a stress. The results were surprising and showed that SAFB2 and not SAFB1 as previously reported was recruited to nSBS. SAFB2 was shown to co-localise with HSF1, SRSF1 and lncSATIII more readily than SAFB1. Moreover, SAFB2 depletion led to an increase in SATIII DNA transcription following recovery after heat stress. These results show that although SAFB1 and 2 are paralogues sharing very similar domains and functions, they still retain unique functions. nSBs are sites where proteins needed for constitutive transcription, splicing and translation temporarily accumulate [192, 209]. Hence, this sequestering of proteins temporarily inhibits the transcription and translation of constitutive proteins [191, 221]. However, as factors needed for the transcription and translation of stress-related proteins are not bound within nSBs the recruitment of proteins must depend on highly specific interactions. As we have found SAFB2 is recruited to nSBs relatively early, it is possible to hypothesise that it may establish interactions with proteins that provide the network to recruit further proteins. It is also possible that SAFB proteins are recruited to the sites of SATIII DNA as co-factors needed for the regulation of SATIII transcription (discussed further below). The increase in SATIII transcription following SAFB2 knockdown suggests it may act with partner proteins to negatively regulate transcription. Although, this could be due to the downregulation of SAFB2 throughout the cell altering the recruitment of other transcription factors to the sites of SATIII transcription (adjacent to nSBs). Further, experiments, for example, investigating the binding of SAFB proteins to SATIII DNA before and after stress could be carried out. This could be carried out using techniques such as chromatin immunoprecipitation (ChIP) with massively parallel DNA sequencing to identify the binding sites of DNA-associated proteins. However, SATIII DNA is found exclusively at pericentromeric regions, and normal sequencing techniques cannot read these highly repetitive GC-rich regions [222]. Therefore, modified PCR based techniques may need to be used. It is also interesting to speculate on the role of post-translational modification in controlling the stress response. For example, is the formation and dissolution of nSBs dependent on proteins being PTM at the initiation and/or end of the stress response? It is possible to speculate that proteins

are post-translationally modified upon stress such that they interact more promiscuously with other transcription factors /RBDs. Upon the end of the stress response, the constitutive PTMs would be restored, and the interactions holding nSBs together lost. These possibilities are investigated further in Chapter 4.

SAFB proteins redistribution following stress

Sat III sequences contain multiple repeated sequences (the GGAAT motif being highly repeated), and HSF1 has been shown to bind directly to Sat III DNA even though the canonical prototypic HSE is not present[107]. Following a stress, non-coding sat III repetitions are transcribed in a polymerase II and HSF1 dependent manner and form the nSB locus [223, 224]. A recent study also confirmed that following a heat shock, HSF1 was required for Sat III expression and that the formation of nSBs was needed for transcriptional repression to occur [225]. nSBs were first characterised as sites of accumulation of heterogeneous ribonucleoprotein (hnRNPs) complexes termed 'peri chromatin granules' (PG), and several proteins (e.g. SRSF1, Sam68 and SRp30c (SRSF9) and CREB binding protein (CBP)) were reported to be recruited to nSBs[223, 224],[107]. . SAFB1 was also reported to be recruited to nSBs, and it was found to co-localise with HSF1 within nSBs after heat stress [1, 11, 74, 78, 191, 202].

Weighardt et al. [1] observed that SAFB formed aggregates (nSBs) under stress conditions. However, the authors used whole SAFB1 purified extracts to generate an anti-SAFB serum (anti-HAP) - SAF-B (4D6): sc-135618, Santa Cruz Biotechnology and this antibody will also recognise SAFB2. Furthermore, the studies of Denegri [74, 191] Rizzi [202], Valgardsdottir [78] and Altmeyer [11] used the same non -specific anti-HAP serum antibody developed by Weighardt [1]. In addition, EGFP tagged SAFB1 deletion mutants were overexpressed, and the location of EGFP was monitored following an HS. These studies showed some SAFB1 is bound to the Sat III/PG loci before a stress; and after a stress, a C-terminal peptide (580-788, within the E/R rich domain) of SAFB1 interacted. The same study also showed that deletions within the 638-788 region of this peptide prevented binding to nSBs. These studies are dependent on constructs being overexpressed, and this method increases the chances of non-specific interactions and is not as reliable as the detection of endogenous proteins.

The comparative images in Figure 3. 31 show that the anti-HAP antibody's recruitment pattern has more similarities with the SAFB2 recruitment pattern seen in our study. Indeed, after 1h of recovery following heat stress, 80% of the large HSF1 nSBs and 55% of lncSATIII nSBs are in co-localisation with SAFB2 nSBs, whereas only 5.5% of the large HSF1 nSB show co-localisation with SAFB1. Our results suggest SAFB2 has unique functions in relation to nSB formation, e.g., it may regulate chromatin opening; SATIII transcription; anchoring lncSATIII to chromatin; recruitment of partner proteins to nSBs. Knocking down SAFB2 was shown to increase SATIII transcription, reflecting an inhibitory role in the early SATIII transcription phase. However, SAFB2 may play a role in a number of these processes, and further research will be needed to understand these roles. These results do suggest unique roles for SAFB2, and observations in non-neuronal studies support this. SAFB2 is highly expressed in the spleen, thymus and hormonally regulated organs, such as the uterus and ovary. SAFB2 expression is also significantly higher than SAFB1 in the male reproductive tract, including prostate, coagulating gland, epididymis and seminal vesicle [31]. SAFB2-/- knockout mice are produced at the expected Mendelian distribution, suggesting SAFB2 is not critical for viability compared to SAFB1-/- where knockout mice are not viable. However, SAFB2 male KO mice displayed a significant increase in several Sertoli cells and testis weight compared to control mice, suggesting a role for SAFB2 on the growth control of Sertoli cells. Adult SAFB2 KO animals also showed reduced expression of the Androgen receptor (AR) [31], which could relate to SAFB2 role as an oestrogen receptor co-repressor [4, 49, 68].

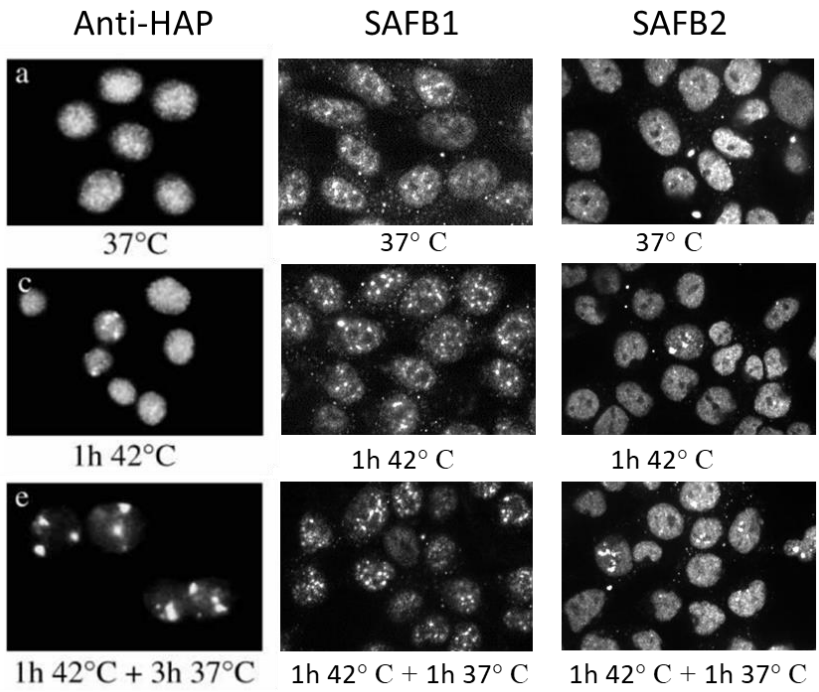

Figure 3. 31 Comparative images between the Anti-HAP Antibody (published by Weighardt, 1999 – left column), SAFB1 and SAFB2 specific staining (middle and right columns respectively). (a) Untreated cells; (c) Cells heat shocked at 42°C for 1 hour; (e) Cells incubated at 37°C for 3 hours following a 1 hour heat shock at 42°C; Image adapted from Weighardt, 1999 (left column [1]).

The results also show that SLTM is is expressed at lower levels than SAFB2,. Although Ally and collaborator's [15] have identified SLTM as a component of a subset of lncSATIII positive nSBs, only qualitative data was presented by the authors, and no quantification was provided. Therefore, the co-localisation between SLTM and lncSATIII positive nSBs shown in Aly`s paper could be represented within the 22% of large SLTM/HSF1 co-localisation seen at the 1Hr HS group (Figure 3. 17 B). Since we have shown that HSF1 nSBs are sites of accumulation of lncSATIII, SLTM might co-localise with HSF1 as well. Ally`s group has shown that the two lncSATIII foci formed after heat stress comprise different subsets of proteins. lncSATIII nSBs containing SAFB proteins (both SAFB1 and SAFB2) also contain SLTM, whereas SLTM is not present in SAFB negative lncSATIII nSBs. This corroborates with the hypothesis that SLTM may associate with either SAFB1 or SAFB1 biding partners during heat stress.

Chapter 04: SAFB1 and SAFB2 Post-Translational Modifications

4.1 Introduction

Post-translational modifications (PTMs) are covalent additions of modifying chemical groups to one or more amino acid residues which alter the functional properties of a protein [226, 227]. Amongst the most common PTMs are phosphorylation, acetylation, glycosylation, amidation, hydroxylation, methylation, ubiquitination and sulfonation. Other less common PTMs include sumoylation, palmitoylation, myristoylation, nitration and Parparylation [228]. SAFB proteins have already been shown to be phosphorylated, acetylated, sumoylated and methylated in several different residues [44, 122, 229-247], modifications which are hypothesised to modify SAFB protein interactions upon stressful conditions. While phosphorylation and acetylation have been well studied across the literature, methylation and sumoylation are less studied but have emerged as important PTMs in recent years.

Methylation is the addition of a methyl group(s) (-CH3) to the n-terminal chain of an amino acid, which could be either an Arginine or Lysine, and it can occur in both histone and non-stone proteins in a reversible process [139, 151, 248, 249]. In histone proteins, methylation is carried out by histone methyltransferases enzymes (HMTs) as part of the histone code, which modulates the chromatin state to guide gene expression [150, 250, 251]. Arginine methylation is catalysed by the Protein Arginine Methyltransferases 1 to 9 (PRMTs 1-9), which generates mono and di-methyl arginine [252]. In contrast, Lysine methylation is catalysed by Lysine Methyltransferases (KMTs), generating mono-, di-, or tri-methyl lysine [139, 142]. Both enzymes promote the transfer of a methyl group from S-adenosylmethionine (SAM/Ado Met) to the guanidino nitrogen in both residues, resulting in the formation of methyl-arginine/lysine and S-adenosylhomocysteine (AdoHcy) [248, 253]. The reverse process, demethylation, is less common, and there are more reports on lysine demethylation than arginine demethylation [155, 254-256]. The added methyl group(s) on arginine's falls into three categories depending on the position of the guanidino nitrogen accepting the methyl group: ω-NG-monomethyl arginine (MMA); ω-NG, NG-asymmetric dimethyl arginine (ADMA); ω-NG, NG--symmetric dimethyl arginine (SDMA). Each addition of a methyl group to an arginine residue removes a potential hydrogen bond donor and changes its shape, altering potential binding partners [152] - PRMTs catalysis type I, II and III reactions. Type I enzymes (PRMT1, 2, 3, 4, 6 and 8) add a methyl group to form MMA and subsequently add a

second methyl group at the same Nitrogen to form ADMA [142] (Figure 4. 1). PRMT 5 and 9 are type II enzymes, and they also produce MMA residues, but the second methyl group is added at a different Nitrogen atom to produce SDMA [257]. PRMT7 is the only known type III enzyme, and it exclusively generates MMA [258].

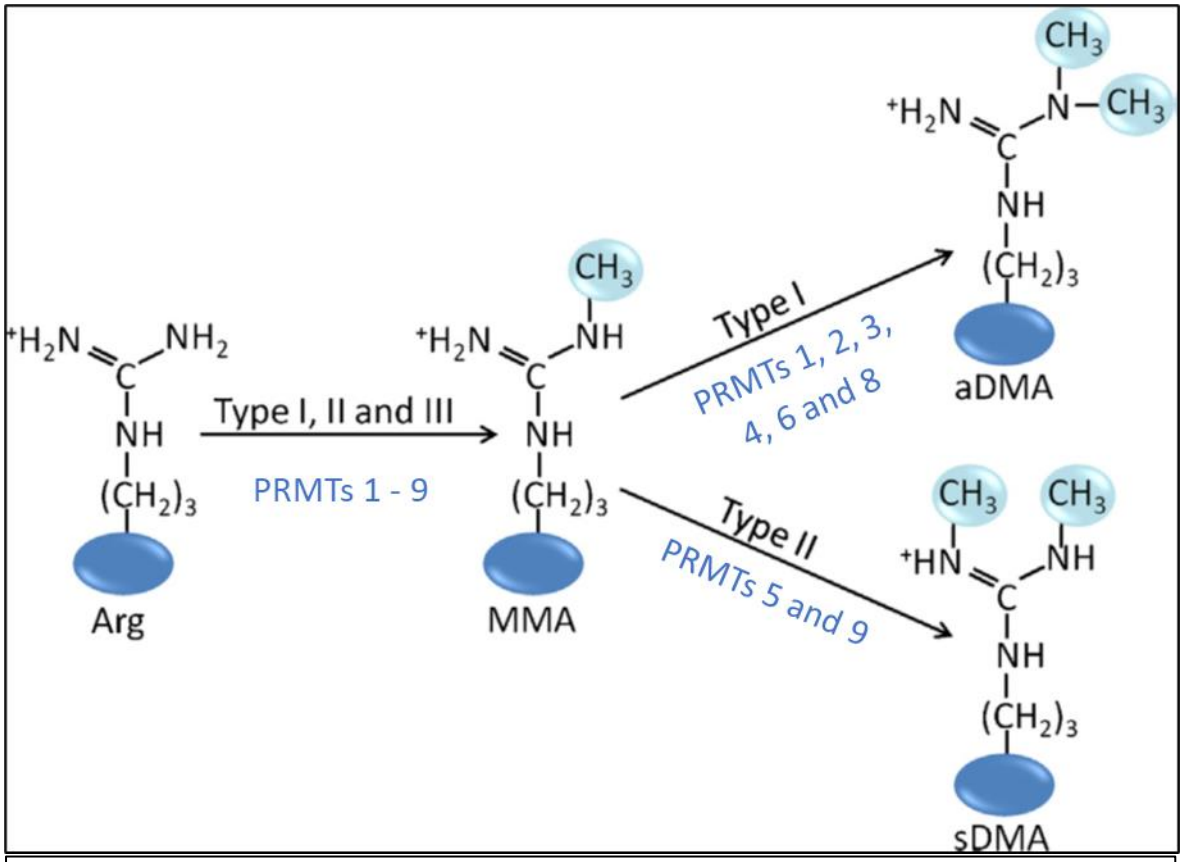

Figure 4. 1 Arginine methylation schematic image. Arginine is mono-methylated by type I, II and III enzymes. Type I enzyme can subsequently add another methyl group at the same amino group, generating asymmetrical di-methylarginine (ADMA), whereas type II enzyme generates symmetrical di-methylarginine (SDMA) by adding the methyl in a separate amino group. Type III enzyme only generates mono-methylarginine (MMA). Adapted from Yannick Auclair and Stéphane Richard, 2013.

Amongst the nine PRMT enzymes, PRMT1 and PRMT5 are the most studied: PRMT1 was the first arginine methyltransferase discovered. It is also the enzyme responsible for much of the protein methylation occurring in cells [259]. PRMT1 is present in the nucleus and cytoplasmic compartments [260], generates MMA and ADMA (Type I reactions), and it recognises arginine's flanked by glycine's residues (RG/RGG motif) [160]. The RG/RGG

motif is a recurrent feature in RNA binding proteins, suggesting that methylation could regulate interactions between RNA and proteins [261]. It has been proposed that methylation of hnRNP molecules by PRMT1 influences their trafficking between cell compartments [262]. Interestingly, the SAF-A (Scaffold Attachment Factor A), a nuclear protein similar to SAFB also containing a SAF-box motif, is methylated by PRMT1 [261, 263]. PRMT1 activity on DNA and RNA damage proteins promotes genome stability and regulates RNA metabolism [264]. MRE11, 53BP1, BRCA1 and Pol β play important roles in the DDR and are all methylated by PRMT1 [168-170, 265-267]. PRMT1 can be successfully inhibited by the adenosine dialdehyde (AdOx, Sigma #A7154), an AdoHcy hydrolase inhibitor that causes intracellular AdoHcy accumulation consequently and feedback inhibition of methylation reactions [268-270]. The S-adenosyl L-homocysteine is a major inhibitor of cellular methyl transferases and prevents methylation of new epitopes [268, 271]. Therefore, AdOx is commonly used as a global methylation inhibitor in bioassays [272-277]. AdOx has been wildly used in global methylation inhibition assays since 1974 [268, 270, 271, 278], and most of the literature reports show an effective final in vitro concentration of 20 μM of AdOx for periods of time ranging between 24-72 hrs [122, 270].

PRMT5 is the major generator of SDMA in type II reactions [279], majority present in the nucleus, but recently discovered in the Golgi compartment [280, 281], with higher expression in muscle (especially heart), and testis [282]. Similarly to PRMT1, PRMT5 also recognises RG/RGG motifs largely present in RNA binding proteins [283]. It has roles in transcriptional repression [284, 285], RNA splicing [286, 287], signal transduction [251, 281] and chromatin organisation [288]. PRMT5 is also implicated in the histone code and gene repression: the H3R2me2s, H3R8me2s, H4R3me2s, and H2AR3me2s methylation marks are generated by it [282, 285, 289]. Studies have shown that elective deletion of PRMT5 in neural stem/progenitor cells (NPCs) reduces methylation of Sm proteins leading to aberrant constitutive splicing, such as in several proteins regulating cell cycle progression, and PRMT5 deletion leads to postnatal death in mice [287]. Moreover, PRMT5 depletion results in the accumulation of DNA:RNA Hybrids and increased spontaneous DNA damage [283]. PRMT5 can be successfully inhibited by the compound GSK3203591 (GSK591), a competitor for PRMT5 substrate with high selectivity against other PRMTs. The GSK591 compound has an IC50 = 4nM and does not inhibit other

PRMTs up to 50 μM. [253, 290]. GSK591 has low toxic effects, and an optimised version of the compound, the GSK3326595, is currently under human trials as an anti-cancer drug [291, 292].

SUMOylation is the PTM where a small ubiquitin-like modifier (SUMO) is covalently bonded to lysine's target proteins [116, 293-295]. The SUMO family is comprised of four confirmed members: SUMO1, SUMO2, SUMO3 and SUMO4 [120, 133, 294, 296-298], and the conjugation of a SUMO modifier to its target proteins requires complex machinery comprised of E1, E2, and E3 enzymes [298-300]. Sumoylation can be reversed by the sentrin/SUMO-specific proteases (SENP), conferring the cell a flexible and rapid way to adapt to external stimuli [135, 301]. SUMOylation has been largely associated with transcription repression [302-305], stress [117, 120, 138, 298] and nuclear integrity [306]. Interestingly, SAFB1 and SAFB2 have been reported to be SUMOylated by the SUMO1 and SUMO2/3 at lysine's 231 and 294, both within the consensus SUMOylation motif ΨKXE, and SAFB1 repression activity depends on SUMOylation. Moreover, the enzyme E3 ligase PIAS1 and SEMP1 are responsible for the modulation of SAFB1`s SUMOylation dynamics [44]. More recently, SAFB1 was shown to be a SUMO-1 substrate bound to the chromatin and associated with RP gene promoters. SAFB1 depletion results in a decrease of RNAPII binding to promoters RNA expression [16]. These results tether transcription initiation and RNA processing of active ribosomal protein genes.

Aims

The post-translational modification of proteins is known to be an important regulatory mechanism. PTMs are known to occur following stress, and these change protein-protein interactions and mediate changes in location and function such that the cell is protected from adverse conditions. We have identified SUMOylation and arginine methylation sites within the SAFB1/2 highly conserved and C-terminal domains, respectively. The aim of the experiments described in this chapter was to investigate how SUMOylation and arginine methylation of SAFB regulates their interactions following a heat shock response.

4.2 Results

4.2.1 Effects of methylation inhibitors on the recruitment of HSF1, SAFB1, SAFB2 and SLTM

Since PTMs can alter interactions between proteins and other biomolecules, we sought to evaluate the effects of methylation inhibition on SAFB protein function. Protein arginine methyltransferases (PRMTs) are capable of recognising arginines flanked by glycines (RG/RGG motif) [307]. The SAFB1 protein has five RG and two RGG motifs, whereas SAFB2 has five RG and three RGG motifs. SLTM on the hand has seven RG and 1 RGG motif, making all family members susceptible to arginine methylation. In fact, previous studies have shown SAFB1 arginine methylation on aa811 (MMA), aa868, aa874, aa884 (ADMA) and one RGG motif, aa868 (ADMA). Meanwhile, SAFB2 has two arginine di-methylate sites within RGG motifs (aa897 and aa903, both MMA) and SLTM has one methylated RG motif within the C-terminal domain [122]. We assessed the importance of methylation on SAFB function by using a global methylation inhibitor (AdOx) and also by inhibiting the most functionally important PRMT enzyme, PRMT5, using a specific SDMA inhibitor (GSK591).

SH-SY5Y cells were incubated either with AdOx, GSK591 or DMSO (dilution vector as control) for 48 and 72 hours, respectively, at 37°C. Following this incubation period, cells were heat-shocked for one hour (termed the HS no recovery (HSNR) period and HS with a one-hour recovery period (HS1hrRecovery)). Controls cells were receiving no HS (NHS). Following treatment, cells were fixed, permeabilised and stained with DAPI (nuclear identification), HSF1, SAFB1, SAFB2 or SLTM antibodies. The experiments were repeated on three independent occasions, with a minimum of two repetitions for each condition. Images were taken at the Perkin Elmer Opera LX system under the same parameters used previously (Chapter 03, section 3.2.6). Images were then forwarded to the modular image analysis through the ImageJ software.

Figure 4. 2 to Figure 4. 5 show the effects of global methylation inhibition and arginine symmetric di-methylation inhibition on the recruitment of HSF1, SAFB1, SAFB2 and SLTM during a stress response. AdOx (mediating the global inhibition of methylation)

significantly reduced the number of cells exhibiting large and small HSF-defined nSBs after heat stress. In contrast, the inhibition of SDMA significantly increased the number of cells exhibiting large and small HSF-defined nSBs after heat stress (Figure 4. 2). This increase was also seen during the recovery period. Surprisingly, AdOx treatment increased the number of cells exhibiting large SAFB1-defined nSBs, but not SAFB2-defined nSBs, following heat stress. Both inhibition treatments showed no effects on SLTM puncta formation.

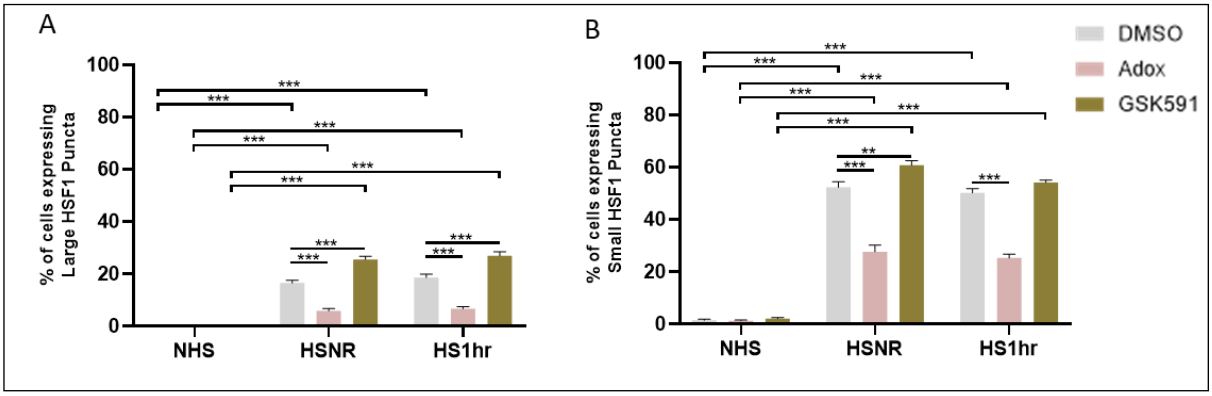

Figure 4. 2 Effects of Global and specific SDMA methylation inhibition by AdOx and GSK591, respectively, on endogenous HSF1. Representative images of HSF1 staining are presented in the following figures (Figures 4.9, 4.10 and 4.11). N=3 independent experiments. Each experiment is comprised by at least 6 wells per condition. A minimum of 10 images per well were included in the analysis. Lines represent mean +/- SEM. *=p<0.05, **=p<0.01, ***=p< 0.001. [= significant difference between time points within the same treatment; | = significant difference between treatments within the same time point.

Representative images of HSF1 staining following treatment with the methylation inhibitors are shown in Figure 4.3 to Figure 4. 5. Treatment with DMSO, AdOx and GSK591 did not prevent the formation of HSF1-defined nSBs, which is shown by the increase in the number of cells exhibiting large HSF1-nSBs following heat stress, in comparison to the NHS controls (DMSO: NHS=0.078±0.02%, HSNR=16.55±1.02% p<0.0001 vs NHS and HS1hrRec=18.58±1.38% p<0.0001 vs NHS; AdOx: NHS=0.12±0.04%, HSNR=5.91±0.9% p<0.0001 vs NHS and HS1hrRec=6.7±0.75% p<0.0001 vs NHS; GSK591: NHS=0.15±0.04%, HSNR=25.54±1.24% p<0.0001 vs NHS

and HS1hrRec=26.98±1.55% p<0.0001 vs NHS - Figure 4. 2, A,). Interestingly, when the inhibition treatments are compared to the DMSO control at corresponding time points, the percentage of cells expressing large HSF1 nSBs immediately after heat stress is decreased in the AdOx treatment (HSNR, p<0.0001 and HS1hr1Rec, p<0.0001 vs DMSO), whereas it is increased by GSK591 treatment (HSNR, p<0.0001 and HS1hr1Rec, p= 0.0004 vs DMSO - Figure 4. 2 A, | bars).

The percentage of cells exhibiting small HSF1 nSBs follows the same trend, showing an increase after HSNR and HS1hrRec (DMSO: NHS=1.48±0.37%, HSNR=52.26±2.18% p<0.0001 vs NHS and HS1hrRec=50.18±1.68% p<0.0001 vs NHS; AdOx: NHS=1.28±0.24%, HSNR=27.68±2.49% p<0.0001 vs NHS and HS1hrRec=25.26±1.5% p<0.0001 vs NHS; GSK591: NHS=2.01±0.42%, HSNR=60.75±1.75% p<0.0001 vs NHS and HS1hrRec=54.22±0.81% p<0.0001 vs NHS - Figure 4. 2 B, [bars). Although the percentage of cells expressing small HSF1 nSBs after heat stress with and without recovery is significantly decreased in the AdOx treated group in comparison to the control group (HSNR, p<0.0001 and HS1hrRec, p<0.0001 vs DMSO - Figure 4. 2 B, | bars), the GSK951 treated group exhibited an increase on its small HSF1 nSBs percentage the immediately after heat stress (p=0.0086 vs DMSO – Figure 4. 2 B, | bar), but not after recovery following heat stress (p=0.0798 vs DMSO - Figure 4. 2 B).

The effectiveness of heat shock treatments is evidenced by the increase of large and small HSF1 nSBs in the DMSO control group (Figure 4. 2). Therefore, heat shock solo effects will not be mentioned unless it is relevant in the context of AdOx and GSK591 treatments.

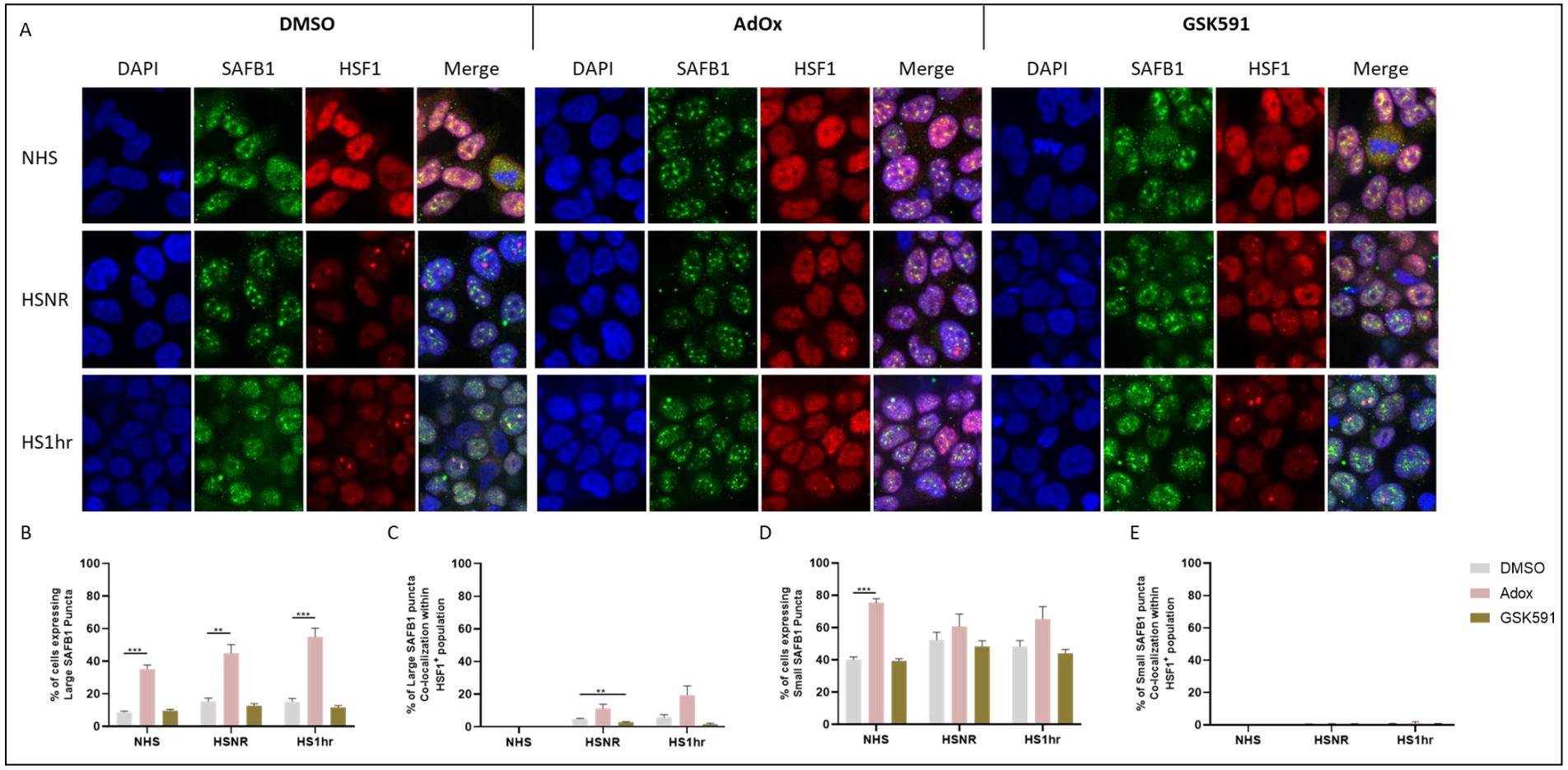

Figure 4.3 Effects of Global and specific SDMA methylation inhibition by AdOx and GSK591, respectively, on endogenous SAFB1 and HSF1. Control images (DMSO) are shown in the left side panel. N=3 independent experiments. Each experiment is comprised by at least 3 wells per condition. A minimum of 10 images per well were included in the analysis per well. Lines represent mean +/- SEM. *=p<0.05, **=p<0.01, ***=p<0.001, |= significant difference between treatments within the same time point.

Following treatment with methylation inhibitors, SAFB1 defined nSBs did show similar changes as seen with HSF1-nSBs (Figure 4.3 B to E). Treatment with AdOx increased the number of cells expressing large SAFB1 puncta in comparison to the DMSO controls under all HS conditions (AdOx: NHS=35.23±2.42%, $p<0.0001$ vs DMSO NHS, HSNR=44.97±5.24%, $p=0.0061$ vs DMSO HSNR, and HS1hrRec=54.95±5.31% $p=0.0007$ vs DMSO HS1hrRec – Figure 4.3B, | bar), whereas there were no significant differences between DMSO and GSK591 treatment at any time point (GSK591: NHS=9.64±0.74%, $p=0.7472$ vs DMSO NHS, HSNR=12.68±1.16%, $p=0.5168$ vs DMSO HSNR, and HS1hrRec=11.66±1.11%, $p=0.2333$ vs DMSO HS1hrRec – Figure 4.3B).

No significant changes could be observed in the percentage of cells displaying large SAFB1 puncta that are co-localised with HSF1 nSBs among the HSF1+ population as a result of a global (AdOx) methylation inhibition in comparison to the DMSO control at all timepoints (DMSO: NHS=0.01±0.01%, HSNR=4.87±0.37% and HS1hrRec=5.76±1.72%; AdOx: NHS=0.03±0.01% $p=0.4482$ vs DMSO NHS, HSNR=11.27±2.62% $p=0.1412$ vs DMSO HSNR and HS1hrRec=19.58±5.48% $p=0.0928$ vs DMSO HS1hrRec – Figure 4.3 C). However, immediately after heat stress (HSNR) the GSK591 treatment significantly decreased this percentage from 4.87±0.37% (DMSO) to 2.78±0.37% (GSK591), with $p=0.0055$ (Figure 4.3 C).

The percentage of cells exhibiting small SAFB1 puncta follows a different trend from the large SAFB1 puncta with no significant changes between time points. Only the AdOx treatment significantly impacted the percentage of cells exhibiting small SAFB1 puncta when corresponding time points are compared across the treatments. However, this is only seen in the absence of stress (DMSO: NHS=40.12±1.70%; AdOx: NHS=75.5±2.38% $p<0.0001$ vs DMSO NHS – Figure 4.3 D). Overall, the results indicate AdOx treatment increases protein-protein interactions (PPIs), resulting in puncta formation.

There are no significant differences in the percentages of cells displaying small SAFB1 puncta in co-localisation with HSF1 nSBs among the HSF1+ population either between time points within treatments or between equal time points across treatments ($p>0.999$ for all comparisons) (Figure 4.3 E).

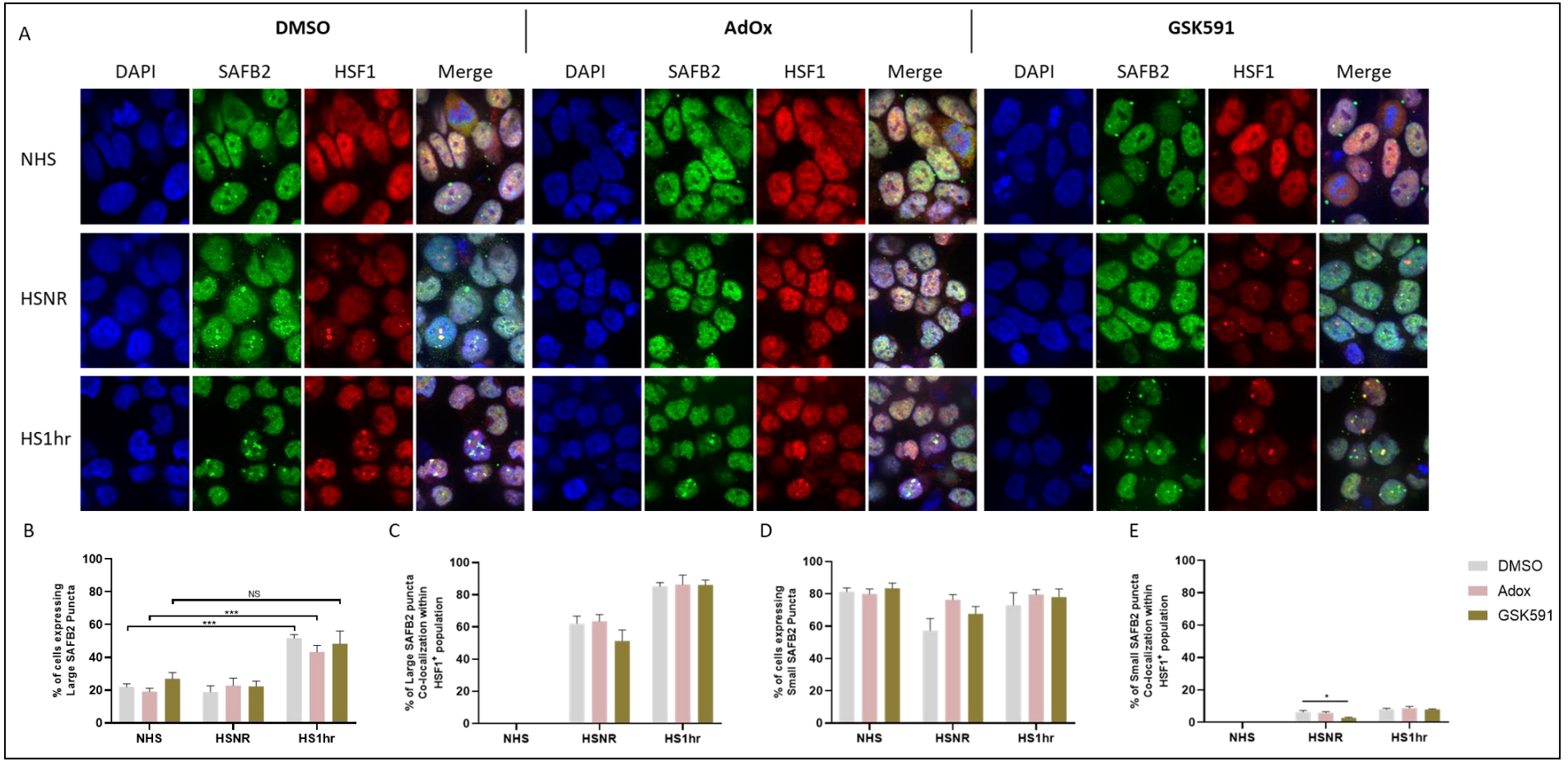

Figure 4. 4 Effects of Global and specific SDMA methylation inhibition by AdOx and GSK591, respectively, on endogenous SAFB2 and HSF1. Control images (DMSO) are shown in the left side panel. N=3 independent experiments. Each experiment is comprised by at least 3 wells per condition. A minimum of 10 images per well were included in the analysis per well. Lines represent mean +/- SEM. *=p<0.05, **=p<0.01, ***=p< 0.001. [= significant difference between time points within the same treatment; | =significant difference between treatments within the same time point.

The effects of methylation inhibition coupled with heat stress on HSF1 and SAFB2 protein expression are shown in figure 4.4. Representative images of HSF1 and SAFB2 immunostaining are also shown in Figure 4. 4 A. The MIA analysis results presented in panel B shows that both AdOX and GSK591 treatments failed in causing a significant effect on SAFB2 puncta formation in comparison to the DMSO control at equal time points. However, while the percentage of cells exhibiting large SAFB2 puncta following heat stress and recovery (HS1hrRec) are significantly increased in the DMSO and AdOx treated groups (DMSO/HS1hrRec=51.66±2.20%, $p<0.0001$ vs NHS; AdOx/ HS1hrRec=43.29±3.96%, $p<0.0001$ vs NHS), the GSK591 treated group does not reach a significant increase when compared to the NHS control (GSK591/ HS1hrRec=48.30±7.82%, $p=0.0837$ vs NHS – Figure 4. 4 B), inferring a delay on the cells ability to form large SAFB2 puncta at this timepoint.

Treatment with the inhibitors also had no effect on the percentage of cells displaying large SAFB2 puncta in co-localisation with HSF1 nSBs among the HSF1+ population, when corresponding time points are compared (AdOx vs DMSO: NHS $p>0.9999$, HSNR $p>0.9999$ and HS1hr1Rec, $p>0.9999$; GSK591 vs DMSO: NHS $p=0.5896$, HSNR $p=0.4363$ and HS1hr1Rec, $p>0.9999$ – Figure 4. 4 C). This percentage increases significantly as expected after heat stress in all treated groups

There were no significant differences in the percentages of cells displaying small SAFB2 puncta between inhibitors treated groups when compared to DMSO at corresponding time points (AdOx vs DMSO: NHS $p>0.9999$, HSNR $p=0.088$ and HS1hr1Rec, $p=0.8857$; GSK591 vs DMSO: NHS $p>0.9999$, HSNR $p=0.5284$ and HS1hr1Rec, $p>0.9999$ – Figure 4. 4 D). Heat stress treatments also did not result in significant changes in the percentage of cells expressing small SAFB2 puncta.

Only GSK591 affected the co-localisation of small SAFB2 and HSF1 nSBs, and a decrease in co-localised puncta was seen following an HSNR (GSK591 vs DMSO: NHS $p>0.9999$, HSNR $p=0.0156$ and HS1hr1Rec, $p>0.9999$ – Figure 4. 4, E | bar). The percentage of cells displaying small SAFB2 puncta in co-localisation with HSF1 nSBs among the HSF1+ population increased significantly after heat stress in all treated groups.

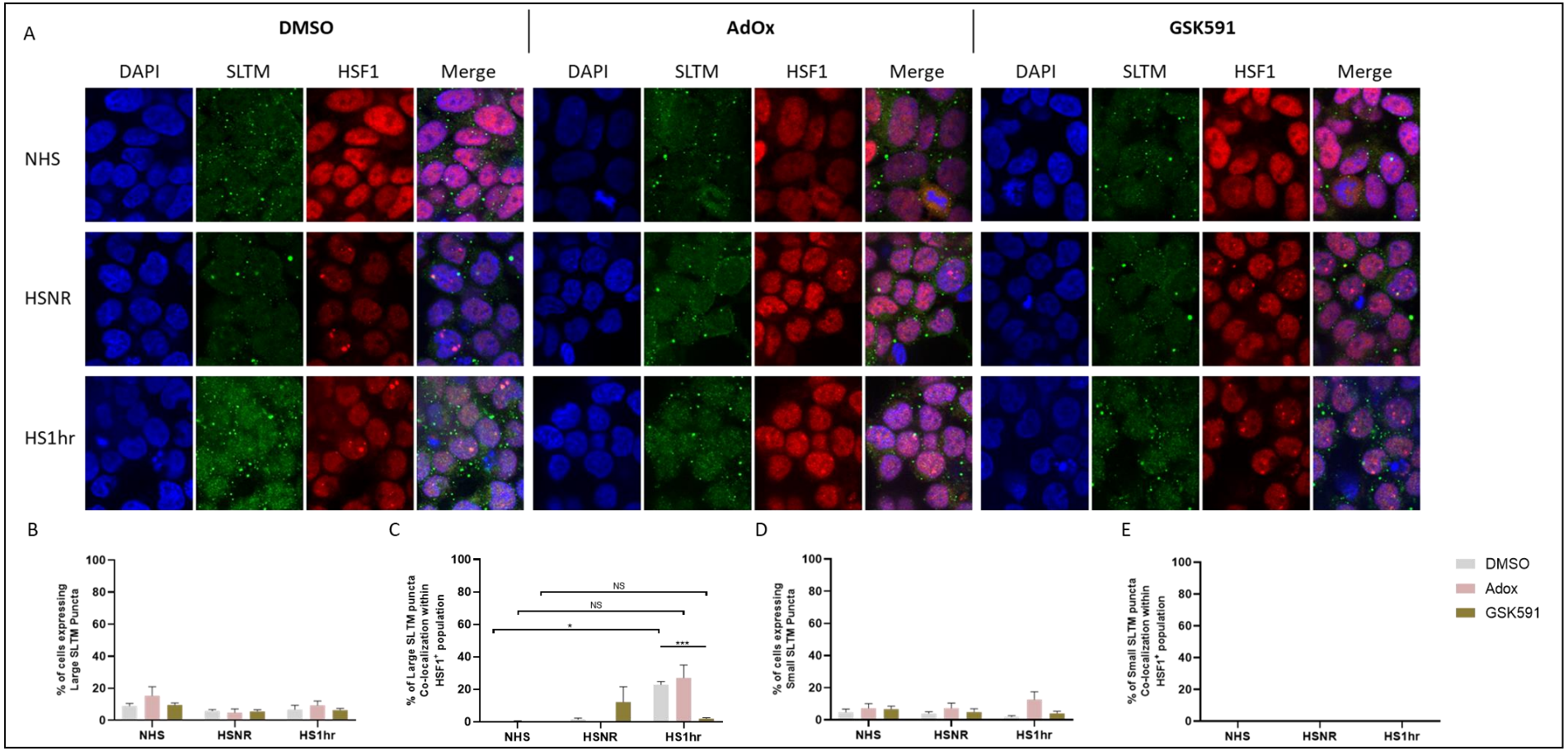

Figure 4. 5 Effects of Global and specific SDMA methylation inhibition by AdOx and GSK591, respectively, on endogenous SLTM and HSF1. Control images (DMSO) are shown in the left side panel. N=3 independent experiments. Each experiment is comprised by at least 2 wells per condition. A minimum of 10 images per well were included in the analysis per well. Lines represent mean +/- SEM. *=p<0.05, **=p<0.01, ***=p< 0.001. [= significant difference between time points within the same treatment; | = significant difference between treatments within the same time point.

Overall methylation inhibitors were found to have no effect on the percentage of cells displaying either large SLTM puncta (AdOx vs DMSO: NHS p=0.6767, HSNR p>0.9999 and HS1hr1Rec, p>0.9999; GSK591 vs DMSO: NHS p>0.9999, HSNR p>0.9999 and HS1hr1Rec p>0.9999 - Figure 4. 5 B) nor small SLTM puncta (AdOx vs DMSO: NHS p>0.9999, HSNR p=0.7336 and HS1hr1Rec p=0.3198; GSK591 vs DMSO: NHS p=0.9875, HSNR p>0.9999 and HS1hr1Rec, p=0.347 - Figure 4. 5 D). Even the heat stress, coupled or not with AdOx or GSK591, was not enough to significantly change the proportion of cells displaying of large SLTM puncta.

There is a significant reduction in the percentage of cells expressing large SLTM puncta in co-localisation with HSF1 nSBs within the HSF1+ population following recovery after heat stress (HS1hrRec) in the GSK591 treated group when compared with the DMSO control group (GSK591 vs DMSO: NHS p>0.9999, HSNR p=0.6788 and HS1hr1Rec p= 0.0108 – Figure 4. 5 C, | bar). The AdOx treatment did not result in similar effect (AdOx vs DMSO: NHS p=0.1638, HSNR p=0.2837 and HS1hr1Rec, p>0.9999 – Figure 4. 5 C). Heat stress only affected the co-localisation of large SLTM puncta with HSF1 nSBs within the HSF1+ population after 1hr of recovery following heat stress (HS1hrRec) in the DMSO group (DMSO: NHS=0.05±0.03%, HSNR=1.50±0.82%, p=0.2559 vs NHS and HS1hrRec=22.84±1.85%, p= 0.0169 vs NHS – Figure 4. 5 C), whereas both methylation inhibitors were able to impair a significant recruitment of SLTM to the HSF1+ nSBs (AdOx: 0.41±0.14% NHS, 0±0% HSNR, p=0.4555 vs NHS and 27.01±7.94% HS1hrRec, p=0.2174 vs NHS; GSK591: 0.03±0.02% NHS, 12.18±9.41% HSNR, p=0.864 vs NHS and 1.98±0.45% HS1hrRec, p=0.0748 vs NHS – Figure 4. 5C). No small SLTM puncta were found to be co-localised with HSF1 nSBs by the MIA analysis (Figure 4. 5 E).

4.2.2 Using deletion mutants to investigate the roles of the SAFB1 functional domains

To investigate the importance of arginine methylation to SAFB protein function we used in vitro mutagenesis to target and substitute the arginine motifs within the SAFB1 and 2 proteins. Expressed of these SAFB deletion mutants was then mediated and their recruitment to $HSF1^{+}$ nSBs compared to wild-type (WT) proteins. Firstly, the expression of SAFB1, 2 and SLTM tagged WT proteins were characterised (methods, section 2.15).

In addition, these analyses investigated the role the functional domains (e.g. the RBD and C-terminal PPI domains) play in SAFB1 protein localisation. Figure 4. 6 (A – C) shows the expression pattern mediated by N-terminal EGFP tagged SAFB1, SAFB2 and SLTM proteins, in HeLa cells, before and after heat stress. Cells were fixed and stained with HSF1 to evaluate co-localisation with SAFB proteins (and recruitment to HSF1-nSBs). Under control (non-HS) conditions, all the SAFB proteins were expressed homogenously throughout the nucleus and were absent from the nucleoli. However, this pattern changes considerably following an HS. Overexpressed SAFB1, SAFB2 and SLTM formed distinct puncta within the nucleus following heat stress. Moreover, SAFB2 puncta also co-localised with HSF1, the primary marker of nuclear stress bodies. This co-localisation was not observed to the same extent with SAFB1 or SLTM proteins. Therefore, the localisation of the overexpressed SAFB1 proteins following HS closely resembles that of endogenous SAFB proteins described above and in chapter 3.

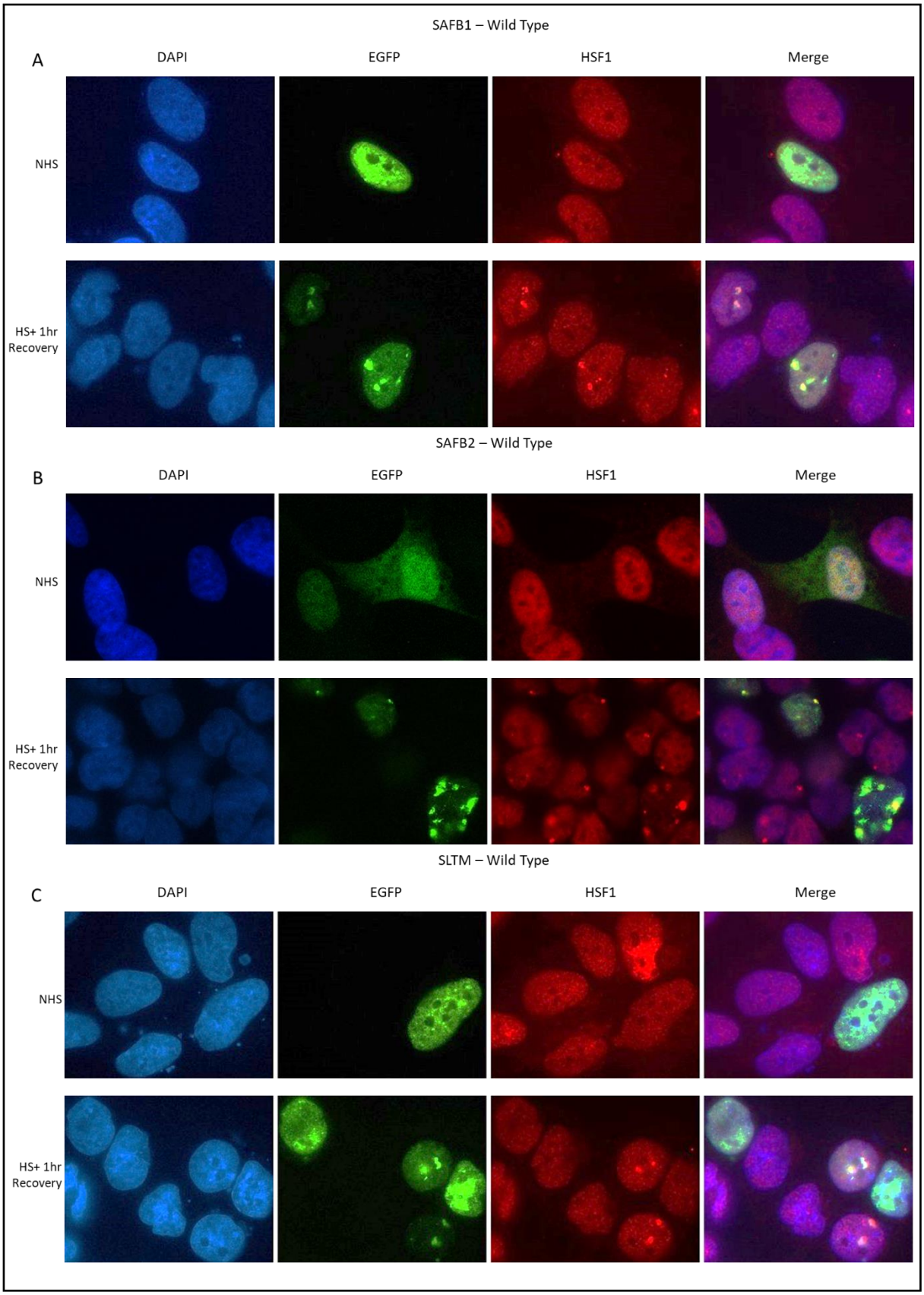

Figure 4. 6 Distribution pattern of overexpressed EGFP tagged Wild Type SAFB1 (A), SAFB2 (B) and SLTM (C), before and after induction of heat stress and recovery. Images co-stained with HSF1 in in red. Oil 63x, 3x zoom.

To analyse SAFB1 domain function, cells were transfected with EGFP tagged SAFB1 proteins lacking either the SAP, HS, RRM, Gly-Rich, R/E-rich and N-terminal portion domains and a wild-type of control (Figure 4. 7)

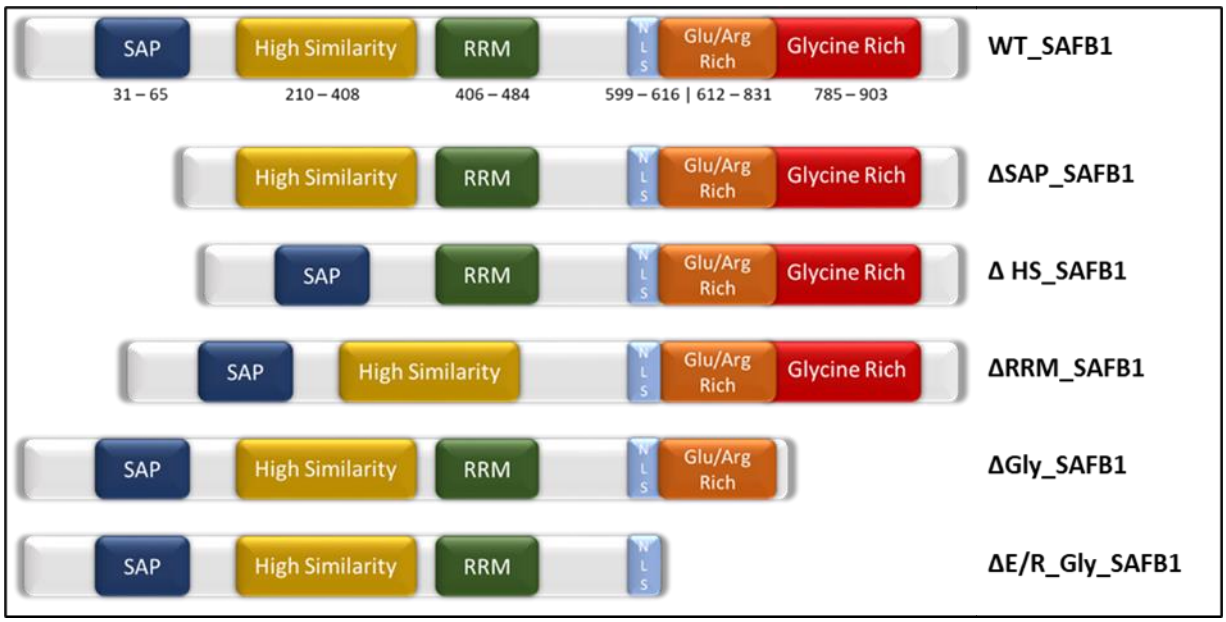

Figure 4. 7 Schematic diagram of the protein sequence in the SAFB1 deletion mutant constructs. Each plasmid encodes a GFP-tagged SAFB1 protein, lacking either the SAF-A/B Acinus and PIAS domain, the High Similarity domain, the RNA Recognition Motif, the Glycine-rich domain, and the N-terminal portion (E/R Gly-rich domains). Adapted from Altmeyer, 2013.

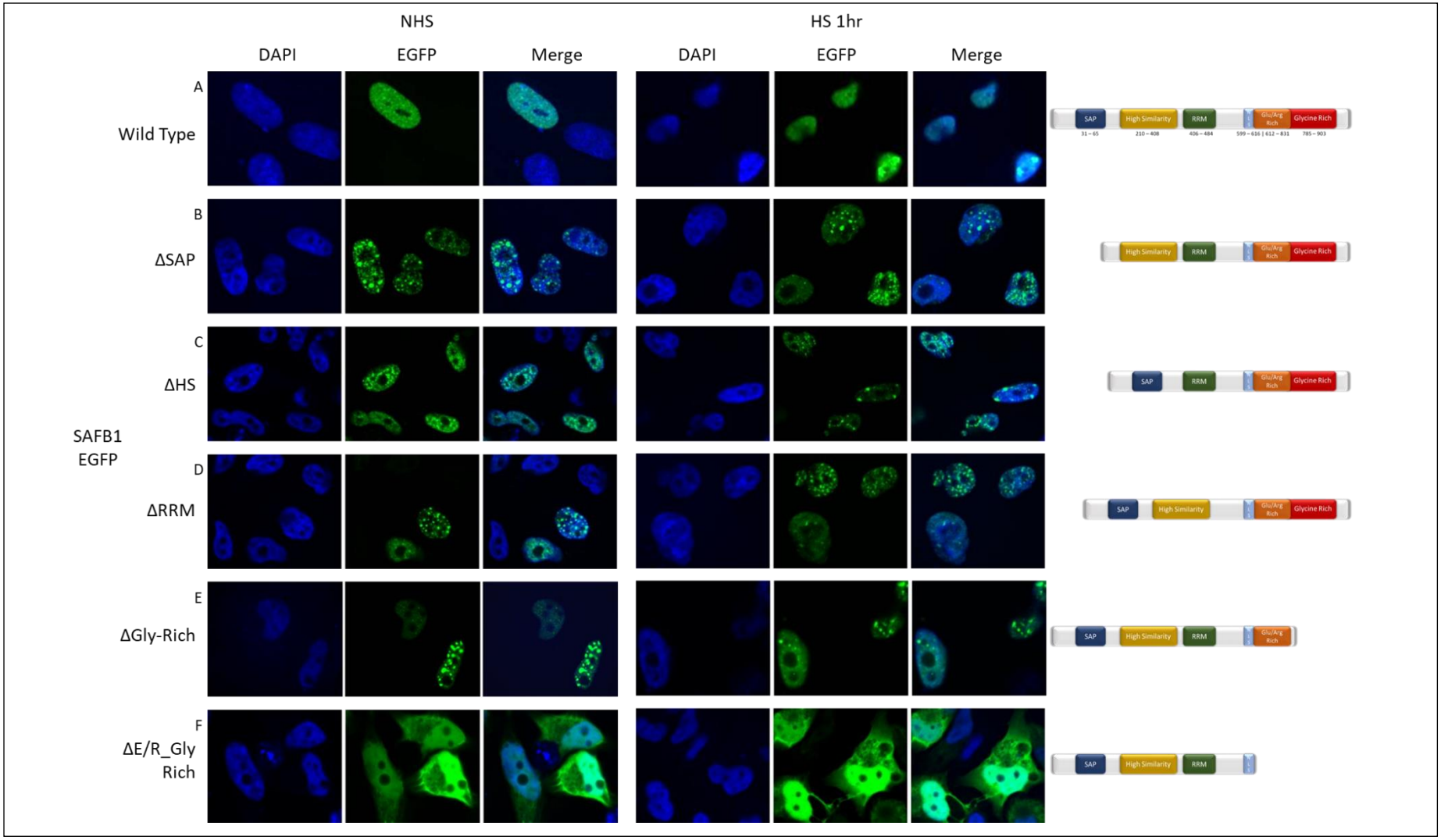

Figure 4. 8 HeLa cells transfected with exogenous EGFP tagged SAFB1 Containing: full length sequence (A), SAP domain deletion (B), high similarity domain deletion (C), RNA recognition motif (D), Glycine-rich domains deletion (E), and E/R-rich domain deletion (C-terminal truncated) (F). Images acquired by confocal laser scanning microscopy with an adaptive focus control, Leica SP8 at 63x magnification.

Immunofluorescent analysis of HeLa cells transfected with an EGFP-tagged WT SAFB1 construct shows a similar SAFB1 expression pattern under control and stress conditions (Figure 4. 8 A) observed in SH-SY5Y cells (Figure 4.3), forming large aggregates following heat stress and recovery. The ΔSAP, ΔHS, ΔRRM and ΔGly mutant proteins are also largely retained within the nuclear compartment, suggesting that these domains are not crucial for SAFB1's nuclear localisation (Figure 4. 9, B, C D and E). Although these domains might not be involved in the protein localisation, they still play a role in the nuclear distribution of SAFB1, as they form small aggregation sites within the nucleus, even in the absence of stress, diverging from the SAFB1 wild type distribution pattern. However, the E/R+Gly-rich domain deletion mutant (F) lost its ability to be retained in the nuclear compartment, showing a homogeneous distribution across the cytoplasm of the cells. This suggests that the C-terminal region of SAFB1 (even though the NLS remains) is needed to maintain SAFB1's nuclear localisation.

Additionally, SH-SY5Y cells were transfected with the ΔE/R_Gly-rich mutant, submitted to heat stress and stained for HSF1 to verify the mutant's ability to be recruited to the HSF1 nSBs (Figure 4. 9). SH-SY5Y cells were also found to express the ΔE/R_Gly-rich mutant in the cytoplasm under control conditions. However, upon stress, the ΔE/R_ mutant moves to the nucleus but do not colocalise with HSF1 nSBs. These results suggest that following a stress, the conformation of the ΔE/R_Gly-rich mutant is altered (possibly by PTMs) such that it interacts with protein(s), which mediate its translocation to the nucleus. Interestingly, the E/R and Gly-rich domains contain 6 out of 7 RG/RGG motifs present in the SAFB1 protein sequence. These RG/RGG motifs can be recognised by PRMTs and are susceptible to Arginine methylation.

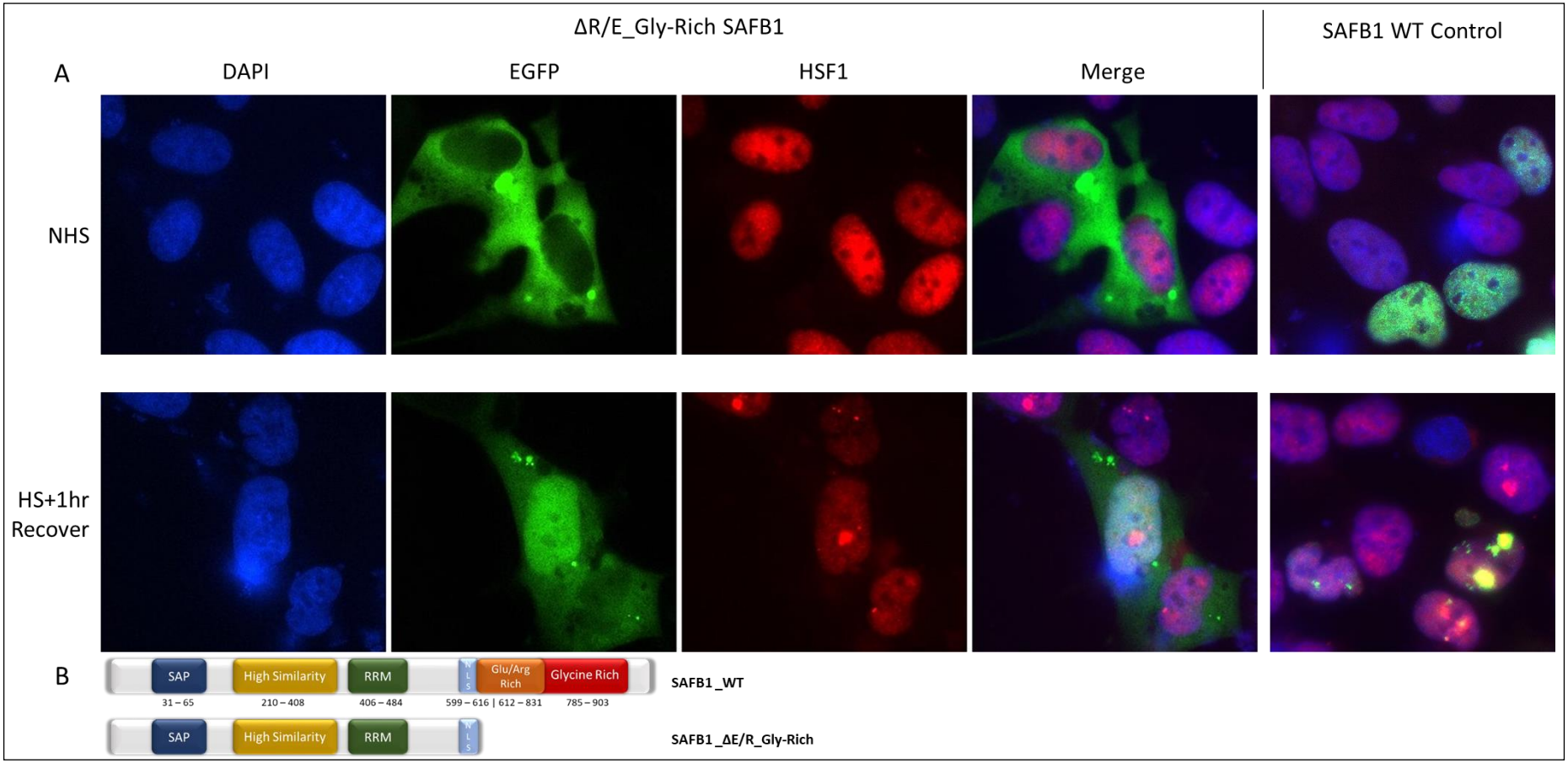

Figure 4. 9 SH-SY5Y cells transfected with exogenous EGFP tagged SAFB1 containing the E/R Gly-rich domain deletion (C-terminal truncated). Images acquired by confocal laser scanning microscopy with an adaptive focus control, Leica SP8 at 63x magnification.

4.2.3 Synthesis of an Arginine x Lysine methylation Mutant of SAFB2

In order to investigate the role arginine methylation may play in SAFB1 and SAFB2 function during stress, the RG and RGG motifs present in both SAFB1 and SAFB2 proteins were mutated (using the Agilent Quick-change II kit, described in methods) such that arginine residues were replaced by lysines generating a SAFB1/SAFB2 KG/KGG mutant. SAFB1 RxK mutations were made by Dr Mamdouh Allahyani, whereas the generation of SAFB1/SAFB2 SUMOylation mutants were made by Dr Nicola Buckner. I designed primers and produced the mutated KG/KGG SAFB2 construct. A pEGFP.SAFB2 plasmid, with a CMV promoter, was used as a template to mutate the arginine methylation sites (aa566, aa773, aa835, aa842, aa883, aa897, aa903 and aa932). More details regarding the generation of these mutants can be found in Methods, section 2.16. A schematic representation of the SAFB2 mutants generated, with the eight RG/RGG motifs presently mutated to KG/KGG motifs (altering the recognition sites for the PRMT enzymes), is shown in Figure 4. 10. This figure also shows the SAFB2 mutated SUMOylation sites. Following the schematic representation of SAFB2 mutants is the sequencing results of the SAFB2 KG/KGG mutant construct, where the inserted point mutations are highlighted in yellow (Figure 4. 11).

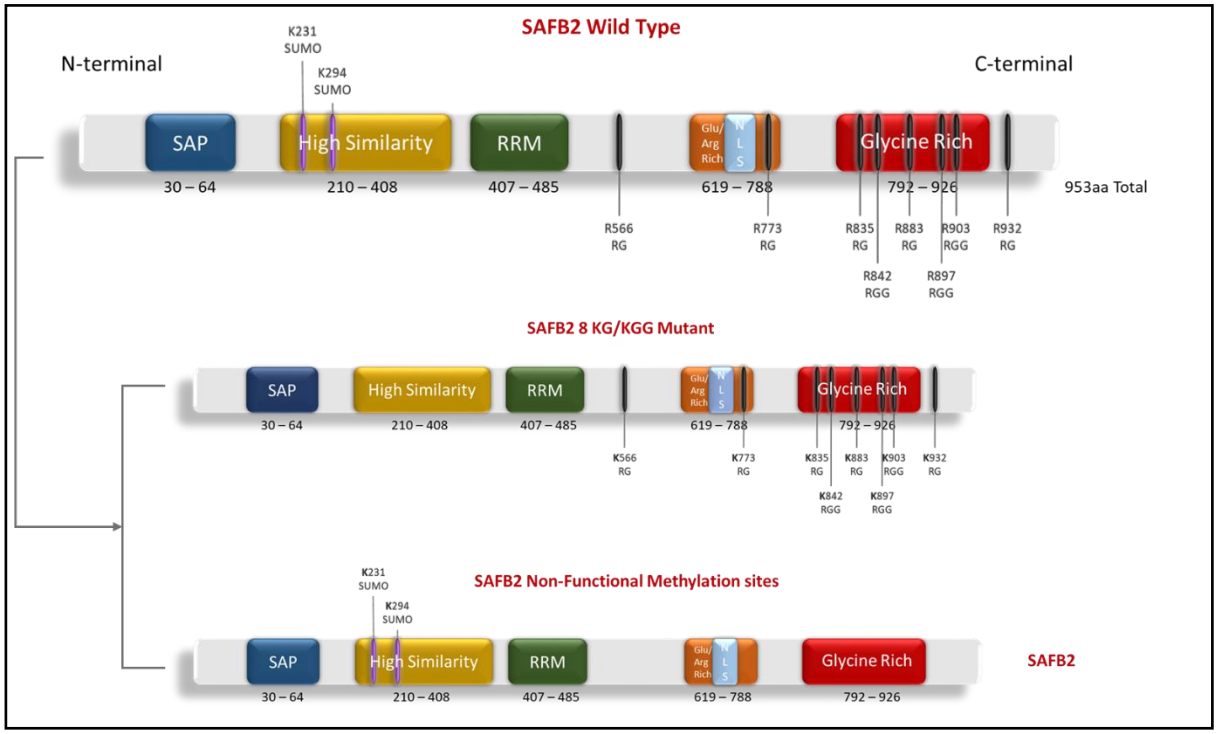

Figure 4. 10 Schematic Image of the wild type SAFB2 protein sequence and its Methylation/SUMOylation motifs (top), followed by the KG/KGG methylation mutants and the non-functional SUMOylation sites on SAFB1/SAFB2 sequences (middle and bottom).

CLUSTAL multiple sequence alignment by MUSCLE (3.8)
Alignment of SAFB2 mRNA sequence and the fully mutated RxK SAFB2 sequence, from aa486 to aa953.

```
SAFB2_WT        GAGCCTGCTGGGAAAAAGCTTTCCGACAGAAAAGAGTGCGAAGTGAAGAAGGAAAAATTA
SAFB2_RxK       GAGCCTGCTGGGAAAAAGCTTTCCGACAGAAAAGAGTGCGAAGTGAAGAAGGAAAAATTA
                ************************************************************

SAFB2_WT        TCGAGTGTCGACAGACATCATTCTGTGGAGATCAAAATTGAAAAAACTGTAATTAAGAAG
SAFB2_RxK       TCGAGTGTCGACAGACATCATTCTGTGGAGATCAAAATTGAAAAAACTGTAATTAAGAAG
                ************************************************************

SAFB2_WT        GAAGAGAAGATTGAGAAGAAGGAGGAAAAAAAGCCTGAAGACATTAAGAAGGAAGAAAAA
SAFB2_RxK       GAAGAGAAGATTGAGAAGAAGGAGGAAAAAAAGCCTGAAGACATTAAGAAGGAAGAAAAA
                ************************************************************

SAFB2_WT        GACCAGGATGAGCTGAAACCCGGACCTACAAATCGGTCTAGAGTCACCAAATCAGGAAGC
SAFB2_RxK       GACCAGGATGAGCTGAAACCCGGACCTACAAATCGGTCTAGAGTCACCAAATCAGGAAGC
                ************************************************************

SAFB2_WT        AGAGGAATGGAGCGGACGGTCGTGATGGATAAATCGAAAGGAGAGCCCGTCATTAGCGTG
SAFB2_RxK       AAAGGAATGGAGCGGACGGTCGTGATGGATAAATCGAAAGGAGAGCCCGTCATTAGCGTG
                *_************************************************************

SAFB2_WT        AAAACCACAAGCAGGTCCAAAGAGAGAAGCTCCAAGAGTCAGGATCGCAAGTCAGAAAGC
SAFB2_RxK       AAAACCACAAGCAGGTCCAAAGAGAGAAGCTCCAAGAGTCAGGATCGCAAGTCAGAAAGC
                ************************************************************

SAFB2_WT        AAAGAAAAGAGAGACATCTTGTCGTTTGATAAAATCAAAGAACAAAGGGAGAGAGAGCGC
SAFB2_RxK       AAAGAAAAGAGAGACATCTTGTCGTTTGATAAAATCAAAGAACAAAGGGAGAGAGAGCGC
                ************************************************************

SAFB2_WT        CAGAGGCAGCGGGAACGGGAGATCCGCGAAACGGAGAGGCGGCGGGAGCGCGAGCAGCGG
SAFB2_RxK       CAGAGGCAGCGGGAACGGGAGATCCGCGAAACGGAGAGGCGGCGGGAGCGCGAGCAGCGG
                ************************************************************

SAFB2_WT        GAGCGGGAGCAACGCCTCGAGGCCTTCCATGAGCGGAAGGAGAAGGCCCGGCTACAGCGG
SAFB2_RxK       GAGCGGGAGCAACGCCTCGAGGCCTTCCATGAGCGGAAGGAGAAGGCCCGGCTACAGCGG
                ************************************************************

SAFB2_WT        GAACGCCTGCAGCTCGAGTGCCAGCGCCAGCGGCTGGAGCGGGAGCGCATGGAGCGGGAG
SAFB2_RxK       GAACGCCTGCAGCTCGAGTGCCAGCGCCAGCGGCTGGAGCGGGAGCGCATGGAGCGGGAG
                ************************************************************

SAFB2_WT        CGGCTGGAGCGCGAGCGCATGCGCGTGGAGCGTGAGCGCAGGAAGGAGCAGGAGCGCATC
SAFB2_RxK       CGGCTGGAGCGCGAGCGCATGCGCGTGGAGCGTGAGCGCAGGAAGGAGCAGGAGCGCATC
                ************************************************************

SAFB2_WT        CACCGCGAGCGCGAGGAGCTGCGGCGCCAGCAGGAGCAGCTGCGTTACGAGCAGGAGCGG
SAFB2_RxK       CACCGCGAGCGCGAGGAGCTGCGGCGCCAGCAGGAGCAGCTGCGTTACGAGCAGGAGCGG
                ************************************************************

SAFB2_WT        CGGCCCGGGCGGAGGCCCTACGACCTGGACCGACGAGATGATGCCTATTGGCCAGAAGGA
SAFB2_RxK       CGGCCCGGGCGGAGGCCCTACGACCTGGACCGACGAGATGATGCCTATTGGCCAGAAGGA
                ************************************************************

SAFB2_WT        AAGCGTGTGGCAATGGAGGACCGATATCGTGCAGACTTTCCCCGGCCAGACCACCGCTTT
SAFB2_RxK       AAGCGTGTGGCAATGGAGGACCGATATCGTGCAGACTTTCCCCGGCCAGACCACCGCTTT
                ************************************************************

SAFB2_WT        CACGACTTCGATCATCGAGACCGGGGCCAGTACCAGGACCACGCCATCGACAGGCGGGAG
SAFB2_RxK       CACGACTTCGATCATCGAGACAAGGGCCAGTACCAGGACCACGCCATCGACAGGCGGGAG
                **********************__**************************************
```

```
SAFB2_WT        GGTTCGAGGCCAATGATGGGAGACCACCGGGATGGGCAGCACTATGGAGATGACCGCCAT
SAFB2_RxK       GGTTCGAGGCCAATGATGGGAGACCACCGGGATGGGCAGCACTATGGAGATGACCGCCAT
                ************************************************************

SAFB2_WT        GGCCACGGAGGACCCCCAGAGCGCCACGGCCGGGACTCCCGTGATGGCTGGGGGGGCTAC
SAFB2_RxK       GGCCACGGAGGACCCCCAGAGCGCCACGGCCGGGACTCCCGTGATGGCTGGGGGGGCTAC
                ************************************************************

SAFB2_WT        GGCTCCGACAAGAGGCTGAGTGAAGGCCGGGGGCTGCCCCCTCCCCCCAGGGGTGGCCGT
SAFB2_RxK       GGCTCCGACAAGAGGCTGAGTGAAGGCAAGGGGCTGCCCCCTCCCCCCAAGGGTGGCCGT
                **************************__******************_**********

SAFB2_WT        GACTGGGGAGAGCACAACCAGCGGCTAGAGGAGCACCAGGCACGCGCCTGGCAGGGTGCC
SAFB2_RxK       GACTGGGGAGAGCACAACCAGCGGCTAGAGGAGCACCAGGCACGCGCCTGGCAGGGTGCC
                ************************************************************

SAFB2_WT        ATGGACGCAGGCGCGGCTAGCCGGGAGCACGCCAGGTGGCAAGGTGGCGAGAGGGGCCTG
SAFB2_RxK       ATGGACGCAGGCGCGGCTAGCCGGGAGCACGCCAGGTGGCAAGGTGGCGAGAAGGGCCTG
                ************************************************_*******

SAFB2_WT        TCTGGGCCCTCGGGGCCGGGGCACATGGCAAGCCGCGGTGGAGTGGCGGGGCGAGGCGGC
SAFB2_RxK       TCTGGGCCCTCGGGGCCGGGGCACATGGCAAGCAAGGGTGGAGTGGCGGGGAAAGGCGGC
                **********************************___***************__*******

SAFB2_WT        TTTGCACAAGGTGGACATTCCCAGGGCCACGTGGTGCCAGGTGGCGGACTGGAAGGTGGC
SAFB2_RxK       TTTGCACAAGGTGGACATTCCCAGGGCCACGTGGTGCCAGGTGGCGGACTGGAAGGTGGC
                ************************************************************

SAFB2_WT        GGAGTGGCCAGCCAGGACCGGGGCAGCAGAGTCCCTCACCCACACCCTCATCCCCCCCCG
SAFB2_RxK       GGAGTGGCCAGCCAGGACAAGGGCAGCAGAGTCCCTCACCCACACCCTCATCCCCCCCCG
                ******************__****************************************

SAFB2_WT        TACCCCCACTTCACCCGCCGCTACTAAGTCCCACTCGCTGCGAGTTTTCGGGTGGGCAGA
SAFB2_RxK       TACCCCCACTTCACCCGCCGCTACTAA------------------------------
                ***************************

SAFB2_WT        CGCACTGTTGAATCTGGTAGCCAGGGTTCCCTCGAACTTGGGGGATCTTTTTAAAAGCAA
SAFB2_RxK       ------------------------------------------------------------


SAFB2_WT        AGTAAATCCTGCCACCATGTTGTAGCTCAATACAATGTGAACTCACTTTTTTTTTTTTTT
SAFB2_RxK       ------------------------------------------------------------


SAFB2_WT        TAAATAAATGTGTTCTTGTTCTGCCATTTTTAAATCAAGGTTTCTGTTAACGAGGCATTC
SAFB2_RxK       ------------------------------------------------------------


SAFB2_WT        CATTTTCCATTAATAAAGTTTACCATTCGCA
SAFB2_RxK       -------------------------------
```

Figure 4. 11 Final sequence of the SAFB2 RxK mutant, aligned with the SAFB2 WT mRNA sequence (human). The highlighted nucleotides indicate where the arginines were converted to lysine.

4.2.4 Recruitment of SAFB1/SAFB2 mutated proteins to HSF1 nSBs following stress

Following the synthesis of the SAFB1 & 2 KG/KGG mutant constructs, plasmids were overexpressed in HeLa cells in parallel with SAFB1 & 2 SUMO-deletion mutants (methods, section 2.5). Each replicate was imaged in at least 5 different random fields. The total number of transfected cells expressing HSF1-defined nSBs, SAFB1 & 2 puncta and SAFB1 & SAFB2 puncta co-localised with HSF1 nSBs were counted.

4.2.4.1 SAFB1 KG/KGG and non-functional SUMO sites

The results (Figure 4. 12) show that SAFB1 lacking RG/RGG methylation sites (Δ-RG-SAFB1) and non-functional SUMO sites (Δ-SUMO-SAFB1) are expressed in the nuclear compartment (A-C). Both SAFB1 KG/KGG and Δ-SUMO SAFB mutants show a significant decrease ($p<0.001$ in both cases, Figure 4. 12 D) in the number of cells with puncta formed immediately following heat stress in comparison to the SAFB1 WT control. There is also a significant difference between the Δ-RG-SAFB1 and Δ-SUMO-SAFB1 mutants ($p=0.0327$ – Figure 4. 12 D). However, there is no difference in the number of puncta following HS plus a one-hour recovery period. This suggests that the arginine methylation and SUMOylation status of SAFB1 modulates the protein-protein or protein-RNA interactions needed for puncta formation.

The proportion of cells exhibiting co-localisation between HSF1 nSBs and SAFB1 puncta is not altered at any time points when compared to the corresponding WT control. Interestingly, the proportion of cells exhibiting co-localisation between HSF1 nSBs and SAFB1 puncta is significantly decreased in the Δ-SUMO-SAFB1 mutant construct when compared to the Δ-RG-SAFB1 mutant construct following heat stress and recovery ($p=0.0247$ Δ-SUMO-SAFB1 vs Δ-RG-SAFB1, Figure 4. 12 E)

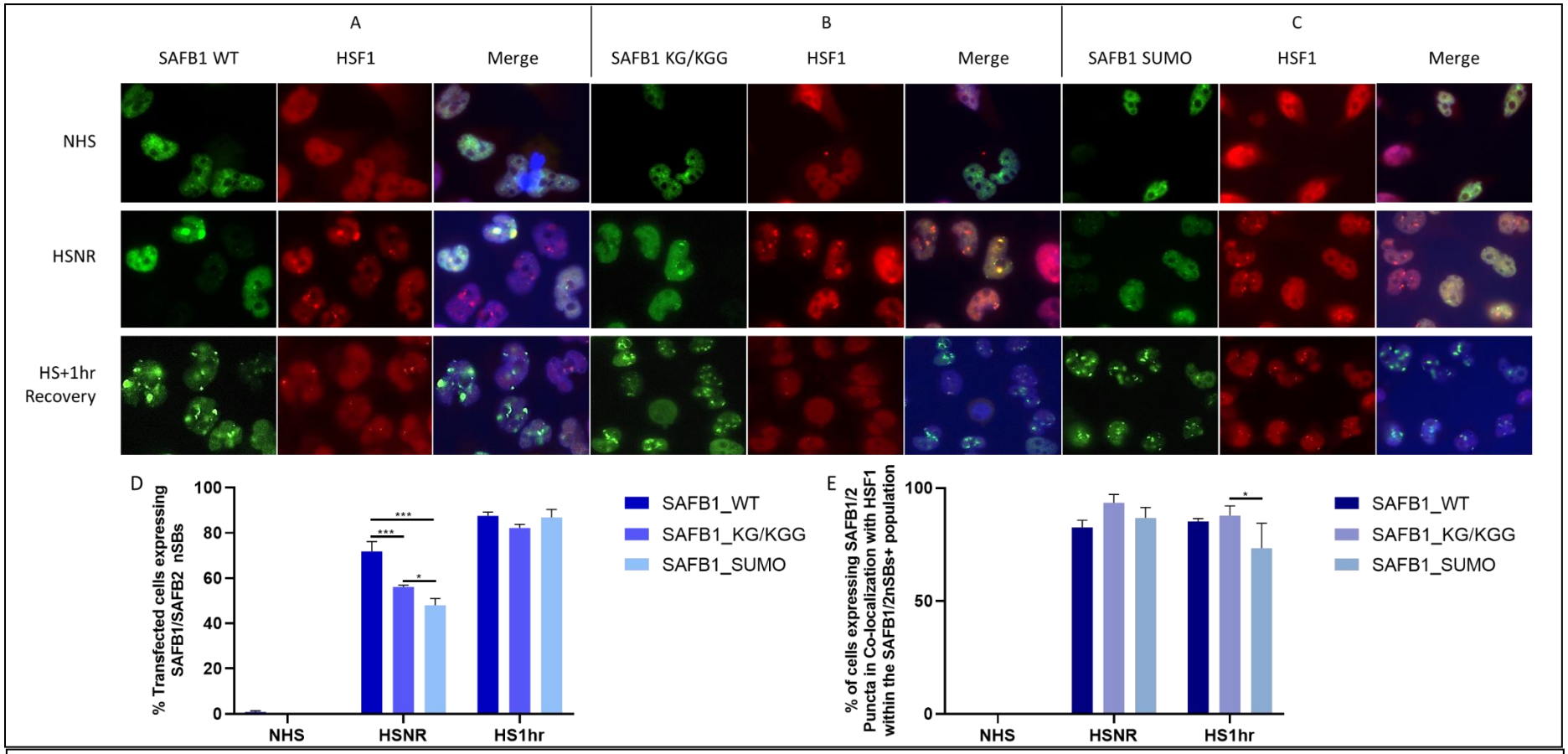

Figure 4. 12 Overexpression of SAFB1 constructs before (NHS) and after one hour of heat shock (HSNR), followed by 1hr of recovery (HS1hr) in HeLa Cells were stained for HSF1. (A) SAFB1 wild type control overexpression(SAFB1_WT), (B) SAFB1 KG/KGG mutant overexpression (SAFB1_KG/KGG), (C) SAFB1 impaired SUMO-sites mutant overexpression (SAFB1_SUMO), (D) percentage of cells expressing SAFB1 puncta, (E) percentage of cells showing co-localization between SAFB1 puncta and HSF1 nSBs within the transfected and HSF1 positive population. N=1 independent experiment comprised by 3 wells per condition. Five images per well were included in the analysis. Lines represent mean +/- SEM. *=p<0.05, **=p<0.01, ***=p< 0.001.

4.2.4.2 SAFB2 KG/KGG and non-functional SUMO sites

Comparative images between SAFB2-WT, Δ-RG-SAFB2 and Δ-SUMO-SAFB2 mutant constructs show no difference in location or size of the SAFB2 puncta in the nucleus (Figure 4. 13 A-C). There was a significant decrease in the number of cells exhibiting Δ-RG-SAFB2 and Δ-SUMO-SAFB2 puncta formed immediately after heat stress compared to the WT control ($p<0.001$ in both cases, Figure 4. 13 D). Interestingly, the number of SAFB2-Δ-SUMO puncta formed after one hour of heat shock followed by recovery was significantly decreased compared to the WT control and the Δ-RG-SAFB2 mutant ($p=0.0006$ vs WT and $p=0.0004$ vs Δ-RG-SAFB2, Figure 4. 13 D). This suggests that the arginine methylation and SUMOylation status of SAFB1 modulates the protein-protein or protein-RNA interactions needed for puncta formation.

Similarly to SAFB1 mutants, the proportion of cells exhibiting co-localisation between HSF1 nSBs and SAFB2 puncta remain unaltered when compared to the corresponding WT control. However, a significant decrease can be seen in the proportion of cells exhibiting colocalising HSF1 nSBs and Δ-SUMO-SAFB2 puncta following heat stress and recovery when compared to the Δ-RG-SAFB2 mutant construct ($p=0.0373$ Δ-SUMO-SAFB2 vs Δ-RG-SAFB1, Figure 4. 13 E).

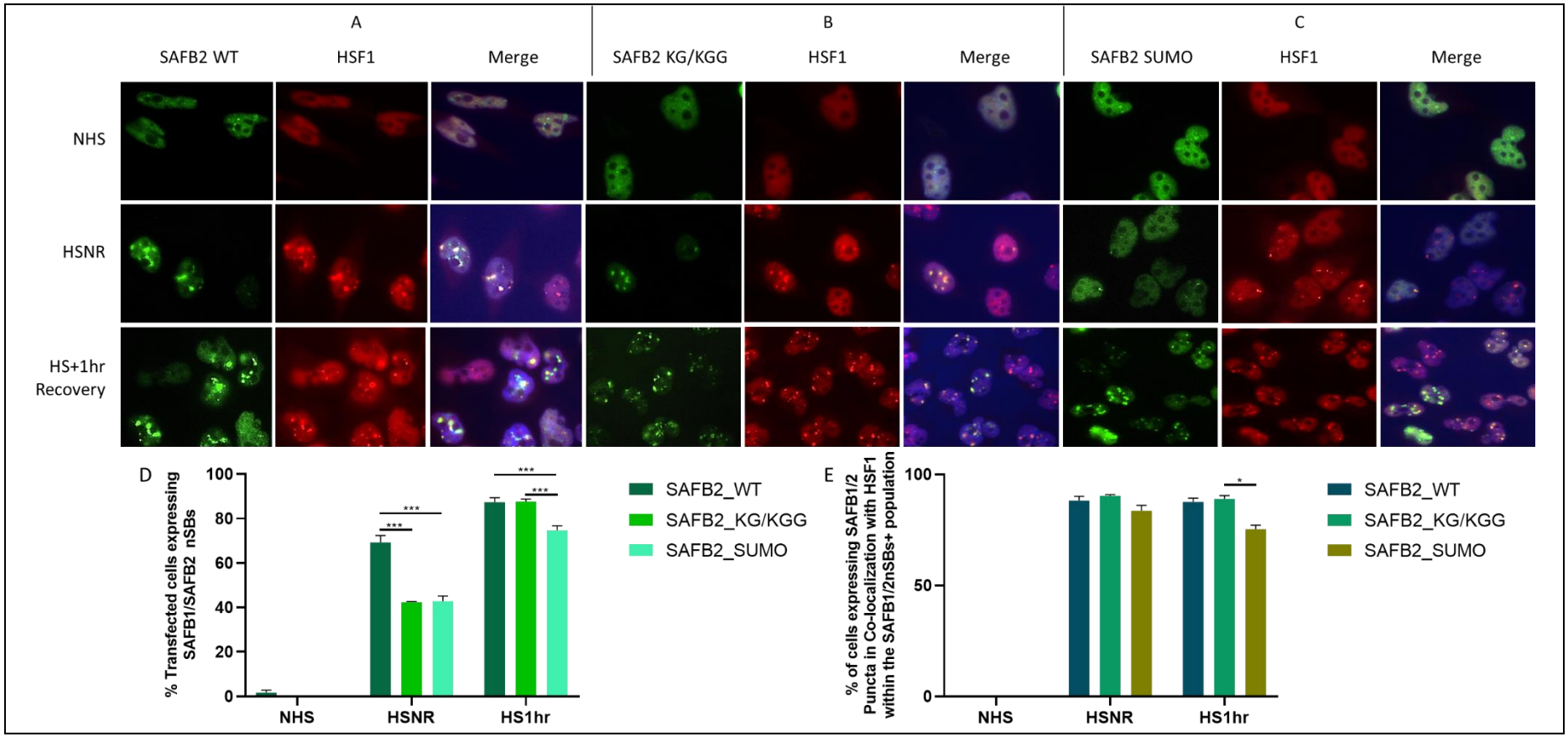

Figure 4. 13 Overexpression of SAFB1 constructs before (NHS) and after one hour of heat shock (HSNR), followed by 1hr of recovery (HS1hr) in HeLa Cells were stained for HSF1. (A) SAFB1 wild type control overexpression (SAFB1_WT), (B) SAFB1 KG/KGG mutant overexpression (SAFB1_KG/KGG), (C) SAFB1 impaired SUMO-sites mutant overexpression (SAFB1_SUMO), (D) percentage of cells expressing SAFB1 puncta, (E) percentage of cells showing co-localization between SAFB1 puncta and HSF1 nSBs within the transfected and HSF1 positive population. N=1 independent experiment comprised by 3 wells per condition. Five images per well were included in the analysis. Lines represent mean +/- SEM. *=p<0.05, **=p<0.01, ***=p< 0.001.

The data from Figure 4. 12 and Figure 4. 13 (percentage of cells showing SAFB1/2 puncta) have been merged (Figure 4. 14) so that the SAFB1 and SAFB2 data can be compared. Immediately after heat stress (HSNR), a lower proportion of cells can form Δ-RG-SAFB2 puncta, in comparison to the Δ-RG-SAFB1 ($p<0.0001$, Figure 4. 14, red arrow). Whereas after recovery following heat stress (HS1hrRec), the SUMO mutants exhibit similar divergent behaviour ($p=0.0024$ Δ-SUMO-SAFB1 vs Δ-SUMO-SAFB2, Figure 4. 14, purple arrow). The graph shows that the PTM of both SAFB1 and SAFB2 by arginine methylation and SUMOylation govern puncta formation following stress.

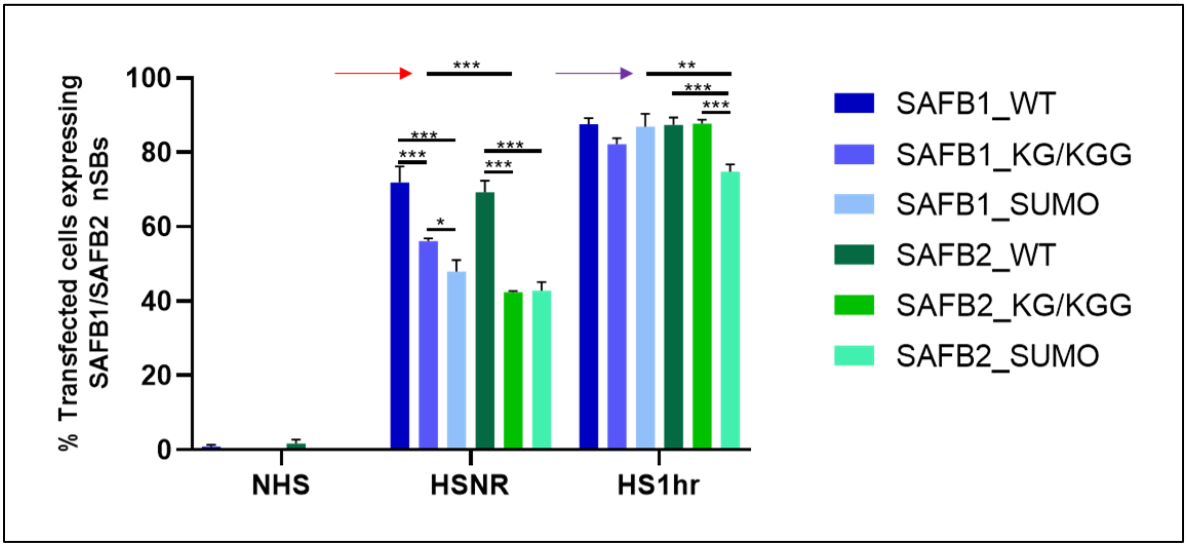

Figure 4. 14 Merged graph of the percentage of cells displaying either SAFB1 or SAFB2 puncta following the overexpression of control and mutant constructs, submitted to various heat treatments. The arrows highlight a significant difference SAFB1 and SAFB2 RG/RGG mutants (red) and SUMO mutants (purple).

4.3 Discussion

In this chapter, the regulation of SAFB1 and SAFB2 by PTMs was investigated. SAFB proteins have been shown to undergo monomethyl arginine (MMA) and asymmetric dimethyl arginine (ADMA) methylation at C-terminal glycine-arginine-rich (GAR) domains (RGG/RG motif) [122]. In addition, two highly conserved SUMOylation sites have been characterised on SAFB1/2 [44]. To investigate whether the methylation of SAFB proteins may alter function, AdOx, a global inhibitor of protein methylation, was used [308]. In the presence of AdOx, the expression of SAFB1 but not SAFB2 was altered, and a greater number of large and small SAFB1 puncta were seen throughout the nucleus. Inhibiting all protein methylation will impact SAFB proteins and its interacting partners. Hence, this change in localisation and aggregation likely reflects a change in binding partner affinity, either due to a methylation mediated change in SAFB1's structure or a change in its binding partner's structure. Cellular stress is known to mediate PTMs. However, no change in the number of SAFB1 puncta was seen following an HS, suggesting that whatever PTM occurs following an HS, they do not counter the effects of global methylation. In addition, global inhibition of protein methylation mediated by AdOX significantly decreased the number of HSF1 large (nSBs) and small puncta formed immediately after and 1 hour after an HS. HSF1 function is altered by methylation. These results suggest that HSF1 binding partner affinity (for proteins and/or SATIII RNA), either due to a methylation-mediated change in HSF1 structure or a change in its binding partner's structure, results in fewer HSF1 puncta/nSBs. There was no change in the expression of constitutively expressed SAFB2 following AdOx suggests that methylation does not regulate SAFB2 function under non-stress conditions. Following a stress, SAFB2 function was not altered compared to controls in the presence of AdOx, and large SAFB2 puncta are seen colocalised with HSF1. This co-localisation in the presence of AdOx suggests that under stress conditions SAFB2 co-localisation with HSF1 is not regulated by methylation or that SAFB2 interacts with the lncRNA that forms the framework for nSB formation. Interestingly, GSK591, a PRMT5 specific inhibitor, was found not to influence the expression of SAFB1 or SAFB2 under control of stress conditions, though it did significantly decrease the localisation of SAFB1 with HSF1. However, GSK591 was found to increase the number of (large and small) HSF1 puncta formed, suggesting PRMT5

mediated symmetrical di-methylarginine (SDMA) residue formation may be necessary for HSF1 interaction. Furthermore, the decrease in the co-localisation of SAFB1 with HSF1 following HS suggests that this interaction may also depend on arginine methylation. SLTM expression was not altered by HS or by AdOx or PRMT5 under control or stress conditions. However, AdOx did increase the co-localisation of SLTM with HSF1 1 hour after HS, suggesting a non-specific interaction occurs when methylation is inhibited. Overall, these results suggest that protein methylation regulates SAFB protein interactions under normal and stress conditions.

SAF proteins contain an SAP-box (a DNA binding motif) that facilitates transcription from scaffold/matrix attachment regions - a C-terminal sequence that mediates protein-protein interactions and an RNA recognition motif (RRM) [3]. There is considerable sequence conservation between SAFB1 and SAFB2 with the SAP-box, RRM, High-Similarity (HS) and C-terminal protein interaction domains sharing greater than 65% homology. To investigate how the SAP, RRM, HS (contains the SUMO domains) and C-terminal PPI domains of SAFB1 may regulate expression, EGFP-tagged deletion mutants were used (Altemeyer et al. and [3]). Deleting the DNA binding SAP domain and RRM altered the pattern of SAFB1 expression, suggesting that SAFB1 interacts directly with DNA and RNA. SAFB1 interacts with many other proteins to help orchestrate the complex interactions of coding, structural and regulatory RNA species [309] (e.g. splicing and DNA repair). Therefore, it is not surprising that deletion of the C-terminal glycine-rich PPI domain also altered the SAFB1 expression pattern. However, deletion of the entire C-terminal PPI domain while retaining the NLS resulted in the EGFP-tagged mutant largely expressed in the cytoplasm. As the NLS remained, this suggests an interaction with another protein is responsible for retaining SAFB1 in the nucleus. Furthermore, following HS expression of this SAFB1-ΔE/R-Gly mutant was seen in the nucleus. This likely indicates that under stress conditions, the PPIs of SAFB1-ΔE/R-Gly protein are altered to translocate to the nucleus. Our laboratory has found that SAFB1 is cleaved by caspases and can be found in the cytoplasm in some cells. This result, coupled with these observations, suggests that SAFB1 may normally shuttle between the nucleus and cytoplasm to carry out some of its functions. Experiments with these deletion mutants

showed that EGFP-tagged overexpressed SAFB1 constructs behaved similarly to endogenous SAFB1 following an HS.

Site-directed mutagenesis was used to modify the arginine methylation and SUMOylation sites of SAFB1 and its expression during HS studied. These experiments showed that the arginine methylation and SUMOylation status of both SAFB1 and SAFB2 regulated their ability to form large puncta immediately following an HS, but their co-localisation with HSF1 was not altered. Furthermore, one hour following this HS period, there was no difference between the puncta formed when compared to controls. This suggests that the SUMOylation and arginine-methylation of SAFB proteins reduce their affinity for their protein (e.g. SRSF1) and/or RNA (SATIII RNA) interaction partners. The RGG domains of the scaffold attachment-A (SAF-A) protein have been shown to play a role in anchoring Xist RNA [310] to DNA. More recently, the differential arginine methylation of RGG motifs (by PRMT1 and PRMT5) of the RNA binding protein, R3BP1, was shown to regulate the assembly of cytoplasmic stress granules, and these results suggest SAFB2 and/or SAFB1 may play similar roles in RNA recognition [311].

Chapter 05: SAFB1 and SAFB2 Interacting Partners

5.1 Introduction

SAFB1 and SAFB2 have a high overall similarity at the amino acid level (74%), suggesting they carry out similar functions [2-4]. Furthermore, SAFB proteins (SAFB1 has been most extensively investigated) have been shown to regulate chromatin organisation [9, 127], gene expression [4, 45, 48, 50, 68], RNA processing and splicing [14, 16, 30, 49, 79, 212], DNA damage and repair [11] and apoptosis [43, 82]. More recently, SAFB1 has been implicated in the biogenesis of polyglutamine disorders, such as Huntington's Disease [95], where abnormal expression of SAFB1 was found in the cytoplasm of striatal neurons in post-mortem brain tissue of HD patients. The same study, conducted by our research group colleagues, also found increased SAFB1 expression in the nucleus and extensive-expression in the cytoplasm of Purkinje cells of spinocerebellar (SCA) ataxia patients, where SAFB1 co-localized with SMI-34, which is a marker for intraneuronal pathological Tau paired helical filaments (PHF-tau/Alzheimer disease tangles) [312] and is a marker of Purkinje cell damage in post-mortem human brain [313].

SAFB1 is primarily nuclear, while SAFB2 is expressed in the nucleus and cytoplasm [4]. In addition, the SAFB protein C-terminal domain is responsible for mediating interactions with proteins that share the least similarity. Together, these data suggest the proteins will interact with distinct partners to modulate similar pathways. The subtle differences between SAFB1 and SAFB2 amino acid sequences may lead to differential PTMs, altering their protein interactions, even changing their interaction partners completely. Therefore, identifying new interactors specific to SAB1 or SAFB2 would help understand their unique functions.

According to the NCBI database, SAFB1 has already been reported to interact with 141 proteins, including SAFB2 [47], HNRNPD [314], RNA polymerase II, SR proteins [127], N-Cor [48], oestrogen Receptor alpha [32, 35, 68], FUS [14], SUMO1, SUMO2/3 [16], HNRNPUL2 [315], ELAVL2, Capn13 [316], SRSF1 [317]. On the other hand, SAFB2 has been reported in the literature to interact with 85 proteins; among them are SAFB1, ESR1, [47], ELAVL2, Capn13 [316], SRSF1 [318], HNRNPA1, Hnrnpa3 and MATR3 [317]. There is considerable overlap between the SAFB1 and SAFB2 binding partners identified;

however, this number may be exaggerated because the antibodies used were raised against SAFB1/2 epitopes and SAFB1 being the most investigated protein.

Aims.

Antibodies were used to distinguish between SAFB1 and SAFB2 in order to:

i. Pulldown SAFB1 and SAFB2 in HeLa cells and use Tandem Mass Tag Mass Spectrometry (TMT-MS) to identify interacting proteins;

ii. Use bioinformatics tools to map the biological processes and functional pathways that SAFB1 and SAFB2 regulate

iii. Highlight the novel pathways regulated by SAFB1 and SAFB2 identified by the proteomic analysis

5.2 Results and Discussion

5.2.1 SAFB1 and SAFB2 interactome isolation and Tandem Mass Tagging spectrometry (TMT-MS)

To identify the interacting partners of SAFB1 and SAB2 we performed co-immunoprecipitation of SAFB1 or SAFB2 followed by TMT-MS (methods 2.8, 2.18). Co-immunoprecipitation of SAFB1 and SAFB2 proteins was carried out in HeLa lysates using protein G magnetic beads cross-linked with either SAFB1, SAFB2 (Bethyl Laboratories) or isotype IgG control antibody (Invitrogen) for background elimination (Methods section 2.17). The lysate was incubated for one hour under constant rotation with IgG antibody cross-linked beads as a pre-clearing step before overnight incubation at 4°C with SAFB1, SAFB2 and control antibodies. The presence of SAFB1 and SAFB2 bands in the western blot pulldown shows the success of the protocol. Dr Mamdouh Allahyani carried out SAFB1/SAFB2 immunoprecipitation (three independent experiments), and details of the protocols used can be accessed in methods section 2.20. Following immunoprecipitation, the samples were processed at the University of Bristol Proteomic facility. Dr Kate Heesom conducted the mass spectrometry steps, including the trypsin denaturation of co-IP samples, TMT labelling, pooling, fractionation, cleaning and analysing by MS. In the

MS final steps, samples were analysed and measured as fold change to the IgG control samples for enrichment analyses, using the Proteome Discoverer Software.

The abundance of SAFB1 and SAFB2 interacting proteins were normalised to the IgG control samples and submitted to a two-tailed T-test on the log-transformed data to generate statistical significance. The mass spectrometry dataset was filtered to include only proteins identified by two or more unique peptides to ensure that the dataset is robust [318]. Interactions with a significantly lower than 0.05 (p-value) in comparison to the IgG control were then selected for further analysis. A Venn diagram (VENNY 2.1 website - http://bioinfogp.cnb.csic.es/tools/venny/) was created with the SAFB1 SAFB2 binding partners (Erro! Fonte de referência não encontrada.). In total, 1072 proteins were identified as SAFB1 binding partners, 288 (24.7%) of them unique to SAFB1, whereas 878 binding partners were found in SAFB2 IP, from which 94 (8.1%) were unique to SAFB2. The remaining 784 proteins were identified as SAFB1 and SAFB2 binding partners, comprising 67.2% of all identified proteins.

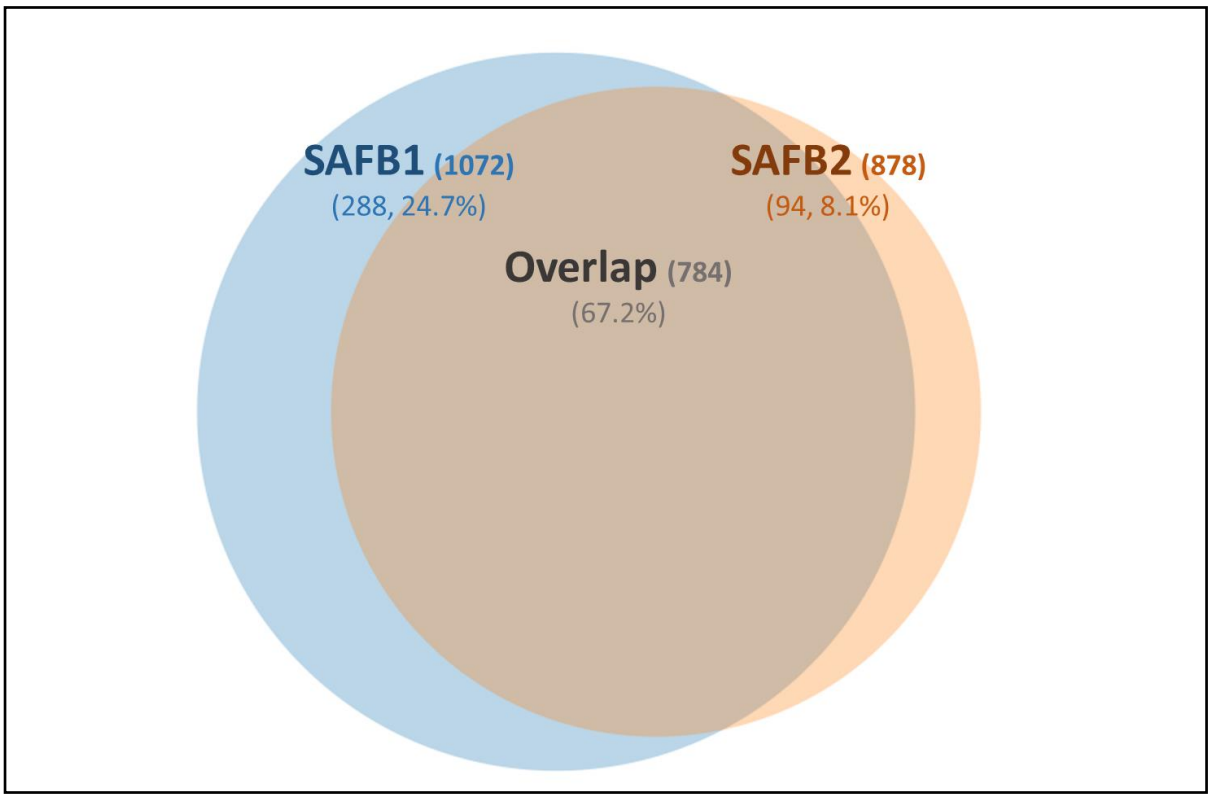

Figure 5. 1 Venn diagram showing the number of SAFB1 and SAFB2 interacting partners and corresponding percentage in HeLa cells. In blue are highlighted the unique SAFB1 interactors, and in yellow are SAFB2 unique interactors. Thus, the shared region comprises SAFB1 and SAFB2 overlapping, interacting biding partners.

The following table (**Erro! Fonte de referência não encontrada.**) shows the top 35 interacting partners for SAFB1 and SAFB2; both are ordered by false discovery rate (FDR). As expected, a large portion of the SAFB1 interacting proteins is RNA binding proteins, with roles in the regulation of gene expression (RBM 26 and 27, ELAVL1, LARP4, NSRP1, FLII), RNA processing (CPSF6, HNRNPA2B1, SYNCRIP, HNRNPR), splicing (EIF4A3, DHX15, SNRNP70), chromatin scaffolding (MATR3, AKAP13) and cell cycle (HP1BP3, CEP170, DHX9). In addition, ELAVL1 [319] and MATR3 [320] are known interactors of SAFB1.

Similarly, the top 40 SAFB2 interaction partners (**Erro! Fonte de referência não encontrada.**) are involved in RNA metabolism, with roles in protein translation (PRKRA, Elongin-A), gene silencing (XPO1), splicing (SNRPD3, EIF2S1, SNRPD2, HNRNPF, RNPS1), RNA export (THOC1, CAPRIN1), RNA processing (HNRNPA0) and RNA binding

protein (NUFIP1). Again, these SAFB2 binding proteins have not been previously reported.

Protein Accession	Gene name	Protein names	SAFB1/IgG		
			LogFCs	T.Test	FDR
Q9P2N5	RBM27	RNA-binding protein 27	5.9306998	4.92182E-05	0.02530356
Q15717	ELAVL1	ELAV-like protein 1 (HuR)	5.2959552	8.33685E-05	0.02530356
Q71RC2	LARP4	La-related protein 4	4.5086203	2.65526E-05	0.02530356
Q9H0G5	NSRP1	Nuclear speckle splicing regulatory protein 1	4.1145924	5.89817E-05	0.02530356
B4DLC0		cDNA FLJ58476, highly similar to Poly(rC)-binding protein 2	2.7230636	7.15045E-05	0.02530356
Q16630	CPSF6	Cleavage and polyadenylation specificity factor subunit 6	7.2629343	0.001172149	0.026747105
O95251	KAT7	Histone acetyltransferase KAT7	6.9304692	0.000301701	0.026747105
Q13045	FLII	Protein flightless-1 homolog	6.3619027	0.000583929	0.026747105
Q12802	AKAP13	A-kinase anchor protein 13 (AKAP-13)	6.2918224	0.00083425	0.026747105
O00425	IGF2BP3	Insulin-like growth factor 2 mRNA-binding protein 3 (IGF2 mRNA-binding protein 3)	6.1627331	0.001036988	0.026747105
Q53F64		Heterogeneous nuclear ribonucleoprotein AB isoform a variant (Fragment)	6.016949	0.000437099	0.026747105
Q9HBB9		HC56	5.8759648	0.00129749	0.026747105
Q9UHI6	DDX20	Probable ATP-dependent RNA helicase DDX20	5.705628	0.00054226	0.026747105
Q9BUZ4	TRAF4	TNF receptor-associated factor 4	5.6341882	0.000895653	0.026747105
Q9NUQ6	SPATS2L	SPATS2-like protein (Stress granule and nucleolar protein)	5.4200117	0.000259827	0.026747105
Q12905	ILF2	Interleukin enhancer-binding factor 2	5.2992152	0.000437671	0.026747105
Q5T8P6	RBM26	RNA-binding protein 26	5.2845695	0.001256985	0.026747105
Q8TCE6	DENND10	DENN domain-containing protein 10	5.2339935	0.0003239	0.026747105
Q08211	DHX9	ATP-dependent RNA helicase A (RNA helicase A)	5.1884463	0.00020494	0.026747105
Q9NZI8	IGF2BP1	Insulin-like growth factor 2 mRNA-binding protein 1 (IGF2 mRNA-binding protein 1)	5.1685177	0.001293588	0.026747105
Q9NNW5	WDR6	WD repeat-containing protein 6	5.1637451	0.001297209	0.026747105
Q5SSJ5	HP1BP3	Heterochromatin protein 1-binding protein 3 (Protein HP1-BP74)	5.1410717	0.001205287	0.026747105
Q92841	DDX17	Probable ATP-dependent RNA helicase DDX17	5.0708511	0.001030306	0.026747105
Q12873	CHD3	Chromodomain-helicase-DNA-binding protein 3 (CHD-3)	5.0673747	0.000908404	0.026747105
P22626	HNRNPA2B1	Heterogeneous nuclear ribonucleoproteins A2/B1 (hnRNP A2/B1)	5.0650778	0.000655537	0.026747105
Q92615	LARP4B	La-related protein 4B	5.0554828	0.001246052	0.026747105
P08621	SNRNP70	U1 small nuclear ribonucleoprotein 70 kDa	5.0389501	0.000526803	0.026747105
Q86XZ4	SPATS2	Spermatogenesis-associated serine-rich protein 2	4.9579946	0.000927374	0.026747105
Q6P158	DHX57	Putative ATP-dependent RNA helicase DHX57	4.9215046	0.000997926	0.026747105
O60506	SYNCRIP	Heterogeneous nuclear ribonucleoprotein Q (hnRNP Q)	4.8894736	0.00044984	0.026747105
A6H8X9	CEP170	Centrosomal protein 170kDa	4.866856	0.001336498	0.026747105
Q5EC54	HNRPK	Heterogeneous nuclear ribonucleoprotein K transcript variant	4.837246	0.000627114	0.026747105
A8MXP9	MATR3	Matrin-3	4.7965545	0.000986378	0.026747105
O43390	HNRNPR	Heterogeneous nuclear ribonucleoprotein R (hnRNP R)	4.7795076	0.000549496	0.026747105
P38919	EIF4A3	Eukaryotic initiation factor 4A-III (eIF-4A-III)	4.7208909	0.001303938	0.026747105

Protein Accession	Gene name	Protein names	SAFB2/IgG		
			LogFCs	T.Test	FDR
Q12834	CDC20	Cell division cycle protein 20 homolog (p55CDC)	3.4536444	0.00011649	0.028804
Q9BVG8	KIFC3	Kinesin-like protein KIFC3	2.1016757	0.000127502	0.028804
Q9UHB9	SRP68	Signal recognition particle subunit SRP68 (SRP68)	2.0827028	0.000133061	0.028804
P58317	ZNF121	Zinc finger protein 121 (Zinc finger protein 20)	2.3374307	0.000101838	0.028804
Q9UHK0	NUFIP1	Nuclear fragile X mental retardation-interacting protein 1	1.839661	6.18697E-05	0.028804
Q9BZH6	WDR11	WD repeat-containing protein 11	5.3719813	8.40651E-05	0.028804
P48643	CCT5	T-complex protein 1 subunit epsilon	1.9031247	0.000126542	0.028804
Q92615	LARP4B	La-related protein 4B	3.9929516	0.000211425	0.0399845
O14980	XPO1	Exportin-1 (Exp1)	3.2500321	0.000474438	0.056325231
P49916	LIG3	DNA ligase 3 (DNA ligase III)	1.7880056	0.000383328	0.056325231
P05198	EIF2S1	Eukaryotic translation initiation factor 2 subunit 1 (eIF-2A)	2.0543969	0.000397931	0.056325231
P68366	TUBA4A	Tubulin alpha-4A chain (Alpha-tubulin 1)	1.7226106	0.000446236	0.056325231
Q96CS3	FAF2	FAS-associated factor 2	0.3698897	0.000482915	0.056325231
Q07666	KHDRBS1	KH domain-containing, RNA-binding, signal transduction-associated protein 1 (Sam68)	3.9225373	0.000574874	0.05978725
Q08379	GOLGA2	Golgin subfamily A member 2	2.6018197	0.000618438	0.05978725
Q658Y4	FAM91A1	Protein FAM91A1	5.0298389	0.000630768	0.05978725
Q14203	DCTN1	Dynactin subunit 1	4.9528864	0.012249958	0.060035328
O95251	KAT7	Histone acetyltransferase KAT7	3.6289889	0.010768316	0.060035328
Q9Y5A6	ZSCAN21	Zinc finger and SCAN domain-containing protein 21	2.9268075	0.003593789	0.060035328
Q96FV9	THOC1	THO complex subunit 1 (Tho1)	4.387861	0.010023999	0.060035328
Q12802	AKAP13	A-kinase anchor protein 13 (AKAP-13)	2.9651449	0.010715379	0.060035328
Q3MHD2	LSM12	Protein LSM12 homolog	4.2350607	0.009615422	0.060035328
O00425	IGF2BP3	Insulin-like growth factor 2 mRNA-binding protein 3 (IGF-II mRNA-binding protein 3)	4.7963296	0.00421834	0.060035328
P62316	SNRPD2	Small nuclear ribonucleoprotein Sm D2 (Sm-D2)	4.0853404	0.013299256	0.060035328
Q53F64		Heterogeneous nuclear ribonucleoprotein AB isoform a variant	4.8236751	0.001697291	0.060035328
Q15287	RNPS1	RNA-binding protein with serine-rich domain 1	3.9417877	0.008858677	0.060035328
Q5BKZ1	ZNF326	DBIRD complex subunit ZNF326	6.3183958	0.003789543	0.060035328
O75569	PRKRA	Interferon-inducible double-stranded RNA-dependent protein kinase activator A	4.6890194	0.009471327	0.060035328
Q9HBB9		HC56	3.067479	0.004058211	0.060035328
Q13151	HNRNPA0	Heterogeneous nuclear ribonucleoprotein A0 (hnRNP A0)	4.4371248	0.011719209	0.060035328
Q9H0M0	WWP1	NEDD4-like E3 ubiquitin-protein ligase WWP1	2.5983246	0.001323263	0.060035328
Q9BRJ6	C7orf50	Uncharacterized protein C7orf50	3.6645297	0.013538766	0.060035328
P62318	SNRPD3	Small nuclear ribonucleoprotein Sm D3 (Sm-D3)	3.8717511	0.013755444	0.060035328
Q14241	ELOA	Elongin-A (EloA) (RNA polymerase II transcription factor SIII subunit A1)	3.0836708	0.011171607	0.060035328

Figure 5. 2 Top 35 proteins identified as associated to SAFB1or SAFB2 amongst all significant interacting partners, classified by false discovery rate (FDR)

Next, we investigated the interacting partners found as unique for either SAFB1 or SAFB2 to highlight the main differences between both proteins. As expected, SAFB1 was shown to interact significantly with more proteins (288 unique binding partners) than SAFB2 (94 unique binding partners).

Among the unique SAFB1 interactors are the Probable ATP-dependent RNA helicase DDX46 (DEAD-box helicase 46) (p=0.005, FDR=0.026), a putative RNA helicase involved in the proliferation cancer cell lines [321-323]; a wide range of E3 Ubiquitin protein ligases (E3 UBLs), such as the E3 UBL Midline-1 (MID1) (p=0.001, FDR=0.026), E3 UBL Itchy homolog (p=0.009, FDR=0.035), E3 UBL ZNF598 (p=0.010, FDR=0.036), NEDD4-like E3 UBL WWP2 (p=0.014, FDR=0.041), E3 UBL RBBP6 (p=0.018, FDR=0.043), E3 UBL MYCBP2 (p=0.033, FDR=0.058), and E3 UBL Jade 2 (p=0.046, FDR=0.067); the E3 SUMO-protein ligases CBX4 and ZNF451(p=0.038, FDR=0.062 and p=0.045, FDR=0.066, respectively); the Sentrin-specific proteases 2 and 3 (p=0.037, FDR=0.062 and p=0.017, FDR=0.043, respectively); the SAFB family member SLTM (p=0.005, FDR=0.032) and catenin alpha-1 (p=0.008, FDR=0.035).

SAFB2 interacts with four subunits of the T-complex protein 1 (TCP) ring complex (TRiC), the TCP subunit epsilon (CCT5) (p<0.001, FDR=0.028), TCP subunit alpha (TCP1) (p=0.003, FDR=0.06), TCP subunit delta (CCT4) (p<0.006, FDR=0.06) and TCP subunit gamma (CCT3) (p=0.035, FDR=0.074). The TRiC is a multi-protein complex chaperone that folds polypeptides [324, 325], responsible for ~10% of the proteome folding [326, 327]. SAFB2 also binds tubulin-alpha4A (TUBA4A) (p=0.0004, FDR=0.0563), a highly conserved homolog of a rat testis-specific protein. It is important to highlight that SAFB2 knockout mice are viable and relatively normal, but they present an abnormal growth of Sertoli cells in the testis [31]. Mutations in the TUBA4A gene have been linked to cases of familial ALS [328]. The FAS associated factor 2 (p=0.0004, FDR=0.0563) is another interesting, unique SAFB2 interacting partner that plays a role in the ubiquitin-dependent degradation of misfolded endoplasmic reticulum proteins [329]. Finally, the Neuroblast differentiation-associated protein AHNAK also interacts exclusively with SAFB2 (p=0.017, FDR=0.06). The AHNAK is a nucleoprotein originally found in epithelial skin cells and neuroblastoma cell lines, with a low expression in the last one [330]. Knockout mice for

the AHNAK gene have been shown to increase cell proliferation and neuronal differentiation into neuroblasts in the hippocampus [331].

5.2.1.1 SAFB proteins involvement in miRNA Processing

The TMT/MS results showed that many of the proteins that interacted with SAFB1 and/or SAFB2 are related to miRNA processing. The results are not entirely unexpected, as previous studies have shown SAFB1/2 involvement in RNA metabolism [49, 127, 332]. However, here we report novel interactions between SAFB1/2 and miRNA processing proteins.

We found SAFB1 and SAFB2 interacted with the Microprocessor Protein Complex (MPC) within the canonical pathway of miRNA biogenesis: Drosha (SAFB1: p=0.012, FDR=0.037; SAFB2: p=0.009, FDR=0.06) and DGCR8 (SAFB1: p=0.032, FDR=0.056; SAFB2: p=0.013, FDR=0.06). DGCR8 is the non-catalytic subunit of the MPC, required for the biding of doubled stranded RNA to the complex, whereas Drosha is a double-stranded RNA ribonuclease III that cleaves the pri-miRNA into a short hairpin structure termed pre-miRNA [333, 334]. The pre-miRNAis exported to the cytoplasm, where Dicer and Argonaute (Ago1-4) will process it to its mature form [335]. Interestingly, SAFB1 and SAFB2 also co-immunoprecipitated with Ago2 (SAFB1: p=0.013, FDR=0.038; SAFB2: p=0.026, FDR=0.067), Exportin-1 (XPO1) (SAFB1: p=0.018, FDR=0.044; SAFB2: p<0.001, FDR=0.056), responsible for the export of Ago2 to the cytoplasm via TNRC6A [336], and TNRC6A (SAFB1: p=0.007 FDR=0.034; SAFB2: p=0.036, FDR=0.06).

Other proteins found to interact with SAFB1/SAFB2 that are also involved in the miRNA processing are MOV10 (SAFB1: p=0.009, FDR=0.036; SAFB2: p=0.014, FDR=0.06), an RNA helicase required for gene silencing by RISC [337]; FMR1 (SAFB1: p=0.014, FDR=0.04; SAFB2: p=0.03, FDR=0.07), which may have roles in mRNA trafficking between nucleus and cytoplasm specially in the brain [338, 339]; ELAVL1 (SAFB1: p<0.001, FDR=0.025; SAFB2: p=0.002, FDR=0.06); PRKRA (SAFB1: p=0.004, FDR=0.032; SAFB2: p=0.009, FDR=0.06); RBM4 (SAFB1: p=0.001, FDR=0.027; SAFB2: p=0.047, FDR=0.08); SRRT (SAFB1: p=0.012, FDR=0.038; SAFB2: p=0.012, FDR=0.06); HNRNPA2B1 (SAFB1: p<0.001, FDR=0.026; SAFB2: p=0.005, FDR=0.06);

eIF6 (SAFB1: p=0.017, FDR=0.043; SAFB2: p=0.025, FDR=0.06); and eIF4E (SAFB1: p=0.016, FDR=0.042; SAFB2: p=0.045, FDR=0.08). **Erro! Fonte de referência não encontrada.** shows the SAFB1/2 interacting partners involved in miRNA processing.

Additionally, SAFB1 and SAFB2 were found to interact with the pre-mRNA-processing factor 6 (PRP6) (SAFB1: p=0.012, FDR=0.04; SAFB2: p=0.034, FDR=0.07), which is involved in the transactivation of the androgen receptor (AR) [340], and it is a component of the U4/U6-U5 tri-snRNP complex, as part of the spliceosome [341]; with the Serine/Arginine-related protein 53 (SRrp53 or RSRC1) (SAFB1: p=0.002, FDR=0.03; SAFB2: p=0.04, FDR=0.08), which plays a role in alternative splicing and transcription regulation [342].

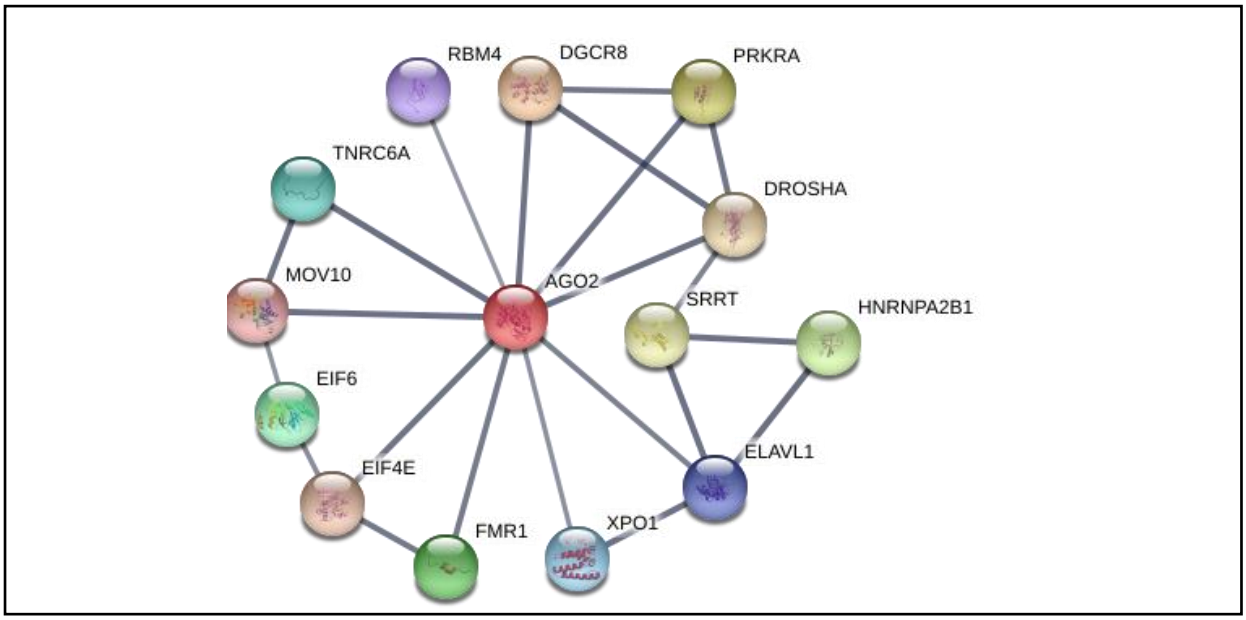

Figure 5. 3 Relevant examples of SAFB1/2 co-immunoprecipitated proteins involved in regulating gene silencing loaded into STRING. Network analyses were based on confidence, whereas the interaction score was set for high confidence (≥0.7). Nodes represent SAFB1/2 interacting proteins.

Interestingly, we found Ago2 - argonaute RISC catalytic component 2 (accession NM_012154) -, amongst the interacting partners of SAFB1 (p=0.013, FDR=0.038) and SAFB2 (p=0.026, FDR=0.067). Ago2 is a member of the RNA-induced silencing complex (RISC) with an important role in small RNAs guided post-transcriptional gene silencing, including mRNA degradation and translational repression (reviewed in [343, 344]), which

has been linked to the development of Huntington's Disease [100]. Our laboratory has an interest in understanding the role of miRNAs in neuronal function and hence this was a particularly interesting finding.

5.2.1.2 Ingenuity Pathway Analyses (IPA)

A wide range of online analyses tools is available to integrate protein data sets [345, 346], such as the Database for Annotation, Visualization and Integrated Discovery (DAVID, updated version 6.8, https://david.ncifcrf.gov/home.jsp) [347], Enrichr (https://amp.pharm.mssm.edu/Enrichr/) [348], WEB-based Gene SeT Analysis Toolkit (WebGestalt - http://webgestalt.org/) [349] and Search Tool for the Retrieval of Interacting Genes (String) version 11.0 (https://string-db.org) [350]. Alternatively, the Ingenuity Pathway Analysis software (IPA - Ingenuity Systems, Silicon Valley, CA) is an all-in-one web-based software application that enables analysis, integration, and understanding of data from a wide range of biological samples such as miRNA, microarrays or proteins. The IPA software calculates P-values using the right-tailed Fishers exact test, which allows the relevant definition of networks. Z-scores were generated in IPA and represent a statistical measure of the match between an expected relationship direction and the observed protein expression. Positive z-scores indicate activation, whereas negative z-scores indicate inhibition. The IPA is powered by the Ingenuity Knowledge Base (IKB) [351], which is manually curated for higher confidence.

Our lists of statistically significant SAFB interacting proteins were imported to the IPA system for pathway analyses. The IKB contains 216 interacting partners for SAFB1 and 145 for SAFB2, which 83 of them are typical for both proteins. Alternatively, the NCBI database shows 141 interacting partners for SAFB1 and 85 for SAFB2, meaning that the IKB holds more evidence for SAFB interactors than the NCBI database. Thus, **Erro! Fonte de referência não encontrada.** shows a Venn diagram comprised of the SAFB1 and SAFB2 interacting partners identified by our experiments and these annotated interacting partners for SAFB1 and SAFB2 within the IKB. Among the SAFB1 interacting partners identified by our experiments, 41 of them are described in the IKB, whereas only 24 are listed for SAFB2 in the database. The remaining identified hits are novel direct/indirect

interacting partners for SAFB1 and SAFB2. A large number of new partners could be explained by the targeted IPs, where specific antibodies against SAFB1 and SAFB2 were used as bait instead of retrieving SAFB1/2 among the results from other IPs.

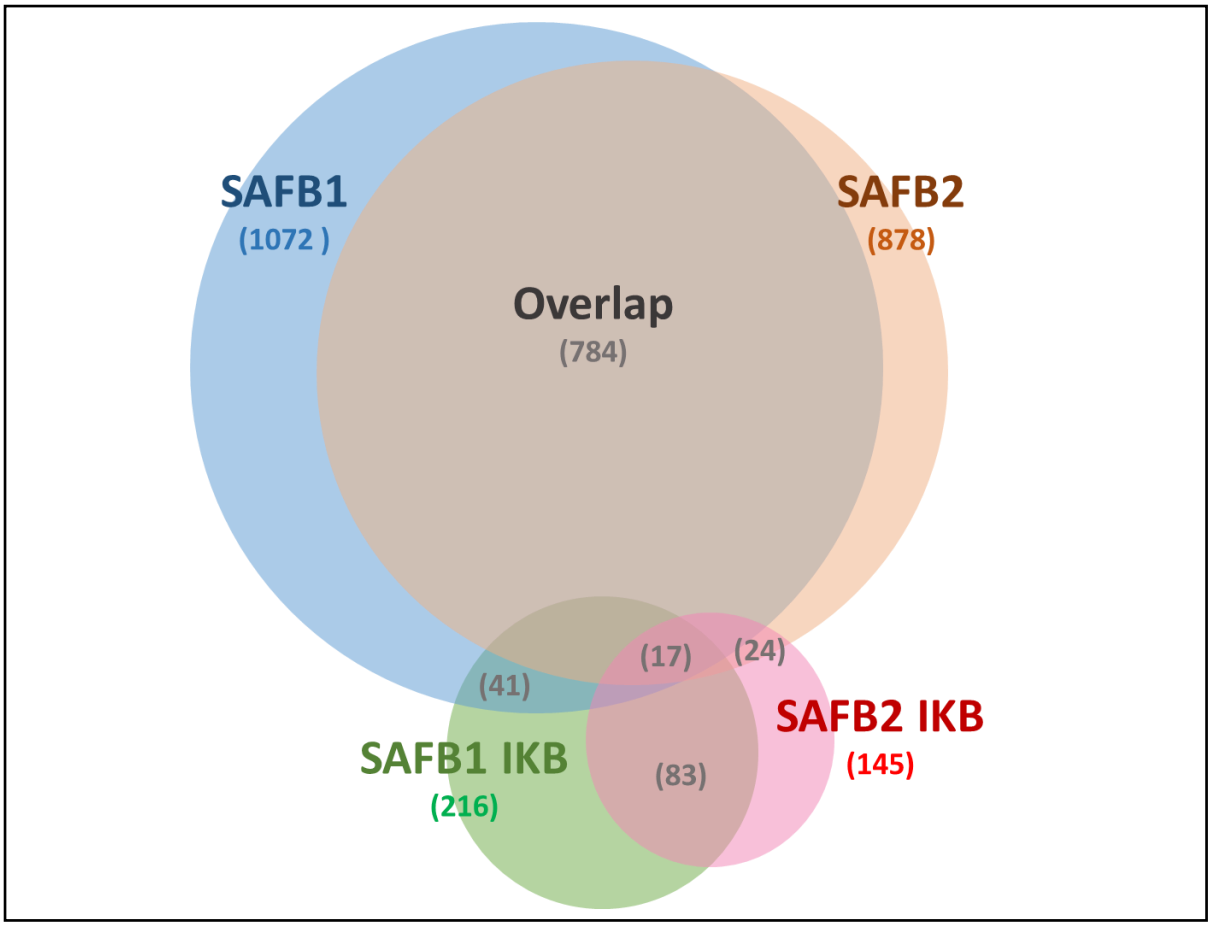

Figure 5. 4 Venn diagram showing the number of SAFB1 and SAFB2 interacting partners identified that are already listed as SAFB1/2 interacting partners in the Ingenuity Knowledge Base (IKB).

5.2.1.3 Canonical pathways enriched for SAFB1 and SAFB2 in HeLa cells.

IPA produced a set of 35 canonical signalling pathways that are significantly enriched for SAFB1 proteins (**Erro! Fonte de referência não encontrada.** A), whereas 41 pathways were significantly enriched for SAFB2 (**Erro! Fonte de referência não encontrada.** B). -Log(p-values)>1.3 are significantly enriched for the protein. In contrast, the Zscore reflects how likely these pathways are activated/inhibited by the protein. Zscores>2 indicates activation, whereas Zscores<-2 indicates inhibition, and in-between Z-scores indicate no effect of the protein on that pathway.

Among the canonical pathways identified enriched for SAFB1 with a positive zScore are the eIF2 signalling (-log(p-value)=51.4, zScore=5.8, Ratio=0.366), mTOR signalling (-log(p-value)=14, zScore=2.12, Ratio=0.19), cell cycle control of chromosomal Replication (-log(p-value) =3.05, zScore=3, Ratio=0.161), Nucleotide Excision Repair Pathway (-log(p-value)= 3.63, zScore=2.88, Ratio=0.136), Unfolded Protein Response (-log(p-value)=1.88, zScore=2.45, Ratio=0.125), and Huntington's Disease Signalling (-log(p-value)=1.63, zScore=2.12, Ratio=0.19) (**Erro! Fonte de referência não encontrada.** A). Although the Role of PKR in Interferon Induction and Antiviral Response pathway ((-log(p-value) =1.39, zScore=-1.89, Ratio=0.08) has a negative zScore, suggesting an inhibition effect on the pathway, it is still not low enough to be statistically significant inhibition.

SAFB2 was found to be enriched among the EIF2 signalling ((-log(p-value) = 38.9, zScore=5.23, Ratio=0.29), mTOR signalling (-log(p-value) = 7.63, zScore=1.13, Ratio=0.13), cell cycle control of chromosomal replication (-log(p-value) = 6.07, zScore=3.46, Ratio=0.21), NER Pathway (-log(p-value) = 5.18, zScore=3.05, Ratio=0.14), Unfolded protein response(-log(p-value)=2.33, zScore=2.23, Ratio=0.13), Huntington's Disease signalling(-log(p-value)=1.48, zScore=N/A, Ratio=0.06) (**Erro! Fonte de referência não encontrada.** B). Interestingly, SAFB2 also appeared to be enriched among the oestrogen Receptor signalling (-log(p-value) = 1.39, zScore=3.77, Ratio=0.05) and tRNA Charging (-log(p-value) = 1.84, zScore=2.23, Ratio=0.13), although SAFB1 was not found among these pathways. The absence of SAFB1 from the

oestrogen Receptor signalling is surprizing, as SAFB proteins (no distinction between SAFB1 and SAFB2) had already been identified as ER coreceptor [47, 48, 68, 352].

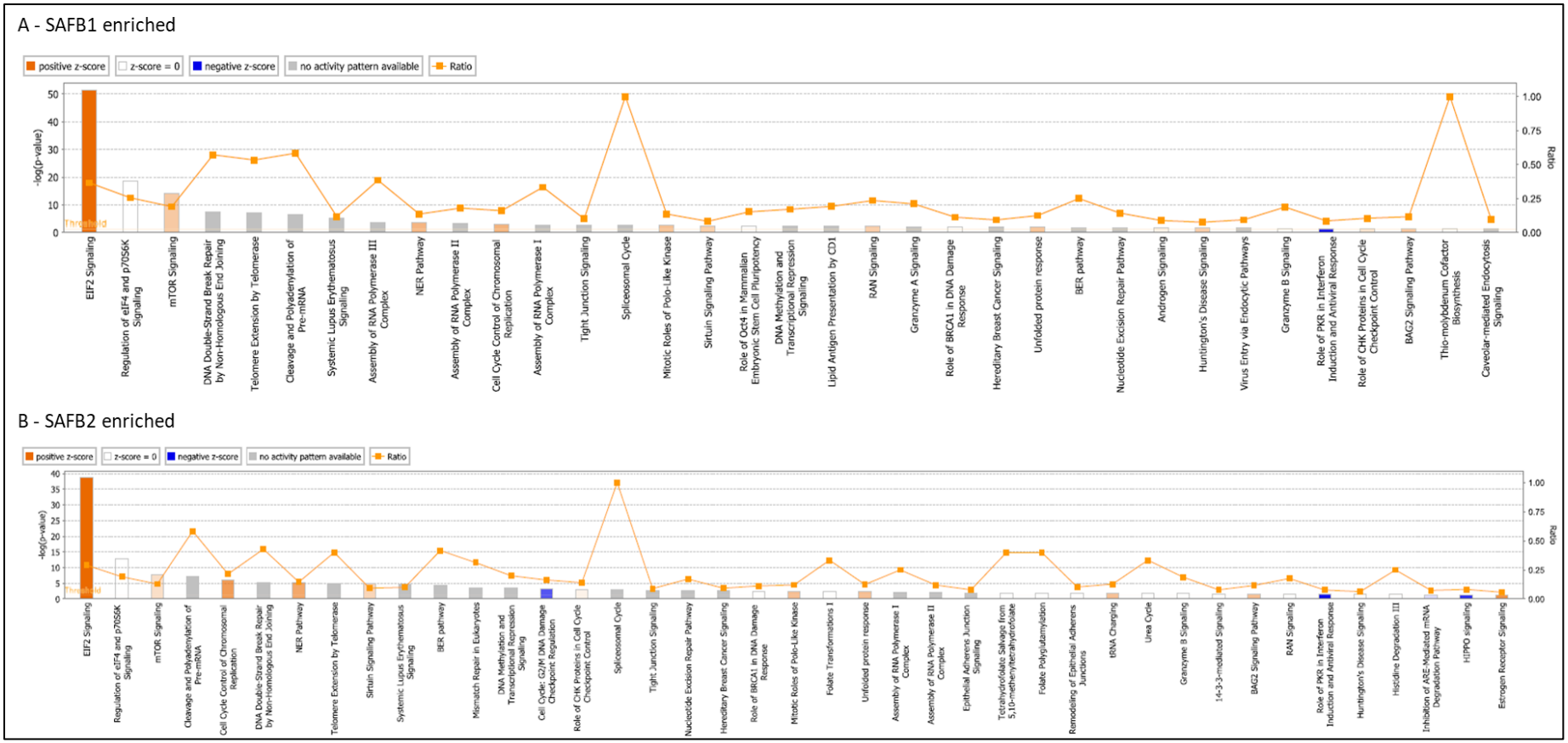

Figure 5. 5 Canonical pathways identified in Ingenuity Pathway Analysis software (IPA) are enriched for SAFB1 and SAFB2 proteins. The canonical pathways identified are based on the 1072 and 878 proteins significantly enriched for SAFB1 and SAFB2 proteins, respectively. Orange refers to a positive z-score, whereas blue represents a negative z-score. Grey represents a pathway with no activity pattern available. The –log (p-value) annotates the most significant changes in pathways.

5.2.1.3.1 SAFB1 and SAFB2 in the eIF2 Signalling pathway

The eIF2 signalling pathway (Eukaryotic Initiation Factor 2) is the most enriched pathway for SAFB1 and SAFB2. Interestingly, this pathway is part of the Integrated Stress Response (ISR), an adaptative pathway to restore homeostasis in response to stress [353], and agrees with previous studies linking SAFB1/2 with the cellular stress response. The pathway presents a -log(p-value) = 51.14 (highly enriched) and a zScore=5.86, suggesting the pathway is more likely to be activated. Eighty-one proteins within this canonical pathway were found to be interacting directly or indirectly with SAFB1, including Ago2 (p=0.013, FDR=0.038) and HNRNPA1 (p=0.003, FDR=0.031), both involved in RNA processing; a variety of 60S ribosomal protein L, such as RPL4, RPL6 to 10, RPL12 to 15, RPL18, RPL21 to 24, RPL26 to 32, RPL35 to 38, and several members of the eIF2 family (EIF2AK2, EIF2S1, EIF2S2, EIF3B, EIF3C, EIF3D, EIF3E, EIF3G, EIF3H, EIF4A3, EIF4E, EIF4G1 and EIF5B). The high number of ribosomal proteins found enriched for SAFB1 corroborates with SAFB1 roles in RNA metabolism. Moreover, the interaction between SAFB1/SAFB2 and members of the eIF2 family links SAFB1 to the ISR. **Erro! Fonte de referência não encontrada.** shows the eIF2 canonical pathway signalling highlighting in magenta the SAFB1 interaction clusters of protein.

Meanwhile, the eIF2 pathway presents a -log(p-value)=38.9 for SAFB2 enrichment ((highly enriched) and a zScore=5.23, also suggesting activation of the pathway. A similar set of interacting partners to SAFB1 is observed here, with Ago2 (p=0.026, FDR=0.66), HNRNPA1 (p=0.025, FDR=0.066), 60S ribosomal protein L (RPL4, RPL6 to 10, RPL12 to 15, RPL18, RPL21 to 24, RPL26 to 28, RPL30, RPL31 and RPL35 to 38), and eIF2 family members (EIF1AX, EIF2B4, EIF2S1, EIF2S2, EIF2S3, EIF3B, EIF3D, EIF3E, EIF3G, EIF3H, EIF4A1, EIF4A3, EIF4E, EIF4G1 and EIF5B) among the SAFB2 interacting partners. Although SAFB2 does not interact with the eIF2α protein kinase PKR like SAFB1, SAFB2 is shown to interact with the PKR-associated protein X (PRKRA), which has been shown to activate PKR in the absence of double-strand RNA **Erro! Fonte de referência não encontrada.** shows the eIF2 canonical pathway signalling highlighting in magenta the SAFB2 interaction clusters of protein.

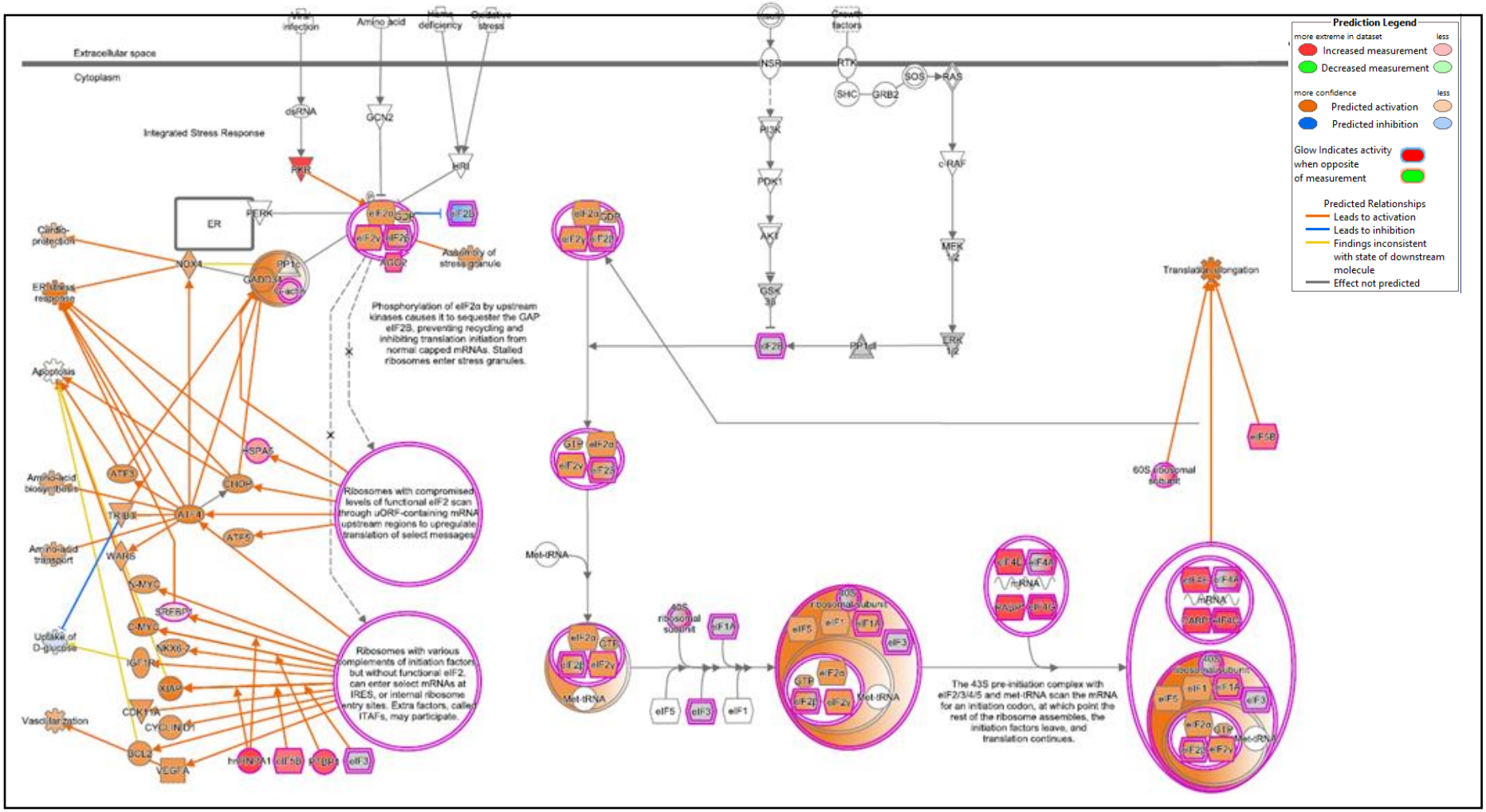

Figure 5. 6 Schematic drawings of the enrichment of SAFB1 within the canonical eIF2 pathway. According to our results, magenta lines represent a protein or group of proteins with which SAFB1 is interacting.

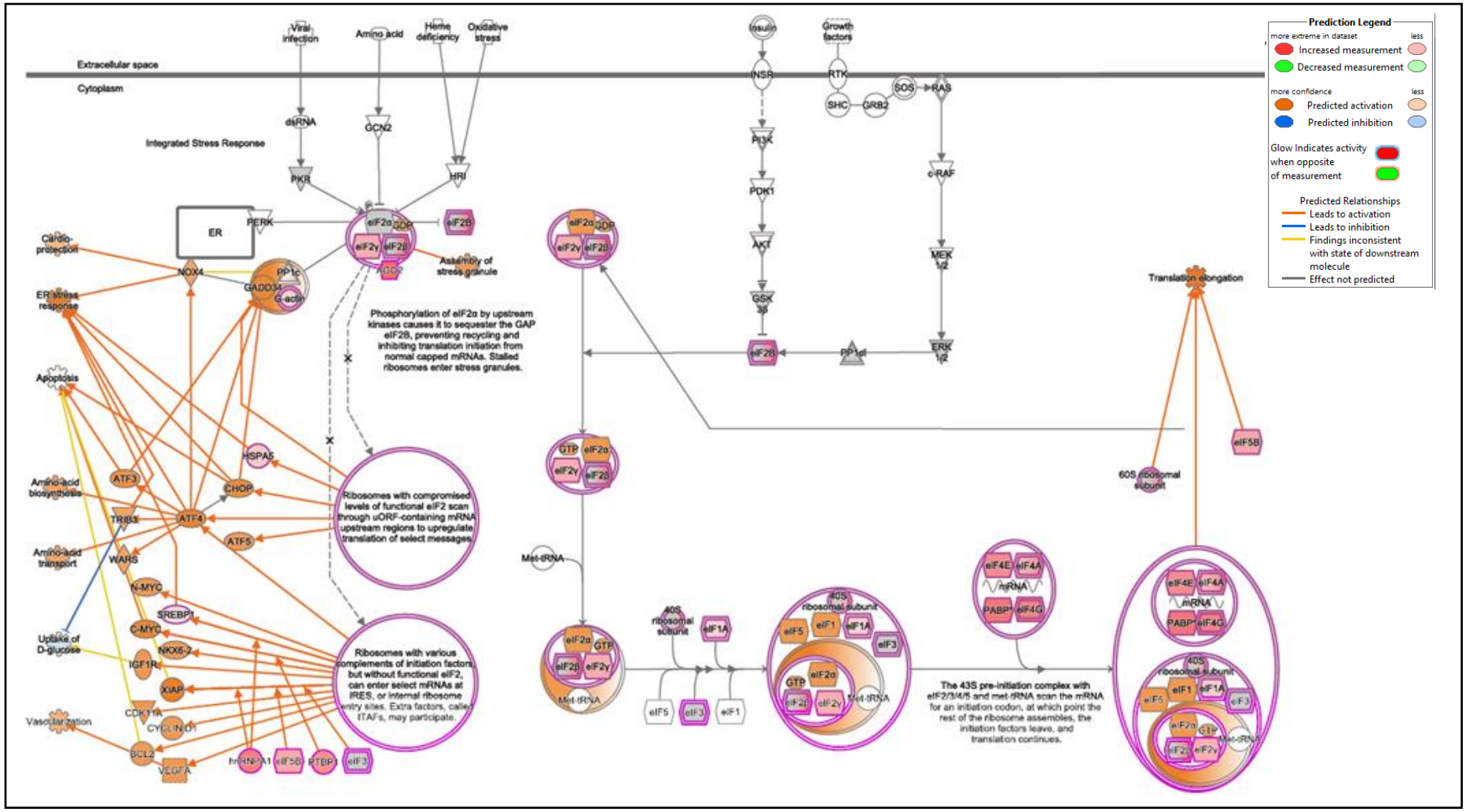

Figure 5. 7 Schematic drawings of the enrichment of SAFB2 within the canonical eIF2 signalling pathway. Magenta lines represent a protein or group of proteins with SAFB2 is interacting with according to our results.

5.2.1.3.2 SAFB1 and SAF2 in the Unfolded Protein Response canonical pathway

The Unfolded Protein Response is also a highly enriched pathway for both SAFB1 and SAFB2 interacting partners. The pathway presents a -log(p-value)=1.88 (significantly enriched) and a zScore=2.45 for SAFB1 and a -log(p-value)=2.33 (significantly enriched) and a zScore=2.23 for SAFB2, in both cases suggesting the pathway is more likely to be activated. **Erro! Fonte de referência não encontrada.** shows the canonical pathway for the UPR and where SAFB1 and SAFB2 proteins interact.

The UPR is a collection of reactions triggered in response to the accumulation of misfolded/unfolded proteins in the endoplasmic reticulum (ER) to return the cell metabolism to homeostasis [354]. It arrests protein production, degrades misfolded proteins and upregulates heat shock proteins to prevent more damage. If these mechanisms are not enough to restore cell function, the UPR leads to apoptosis [355]. Among various causes of stress, heat stress is particularly harmful for the cell proteome as it results in the denaturation of proteins, leading to alteration of its 3D conformations and loss of function or misfunction. SAFB1/2 proteins are known to relocate and form stress granules in the nucleus in response to heat stress [1, 191, 202].

Within the UPR, SAFB1 was found to interact with seven important proteins, five of them are heat shock proteins: calreticulin (CALR, p=0.01, FDR=0.036), HSP40 (p=0.032, FDR=0.058), HSP70 family protein 1B (p=0.03, FDR=0.055) , BIP (HSP70 family protein 5, p=0.046, FDR=0.067), HSP70 family protein 8 (p=0.036 FDR=0.060), HSP70 family protein 9 (p=0.036, FDR=0.061) and SREBF1 (p=0.023, FDR=0.047).

More importantly, BIP can lead to the activation of eIF2α, resulting in ATF4 activation in the nucleus, which leads to the upregulation of BIP and CALR. This action constitutes a link between the UPR and eIF2 signalling.

Similarly, SAFB2 was found to interact with the same five heat shock proteins: HSP40 (p=0.007, FDR=0.06), HSP70 family protein 1B (p=0.007, FDR=0.060) , BIP (HSP70 family protein 5, p=0.003, FDR=0.06), HSP70 family protein 8 (p=0.012 FDR=0.060), HSP70 family protein 9 (p=0.007, FDR=0.06), as well as SREBF1 (p=0.026, FDR=0.067) and VCP (p=0.005, FDR=0.06).

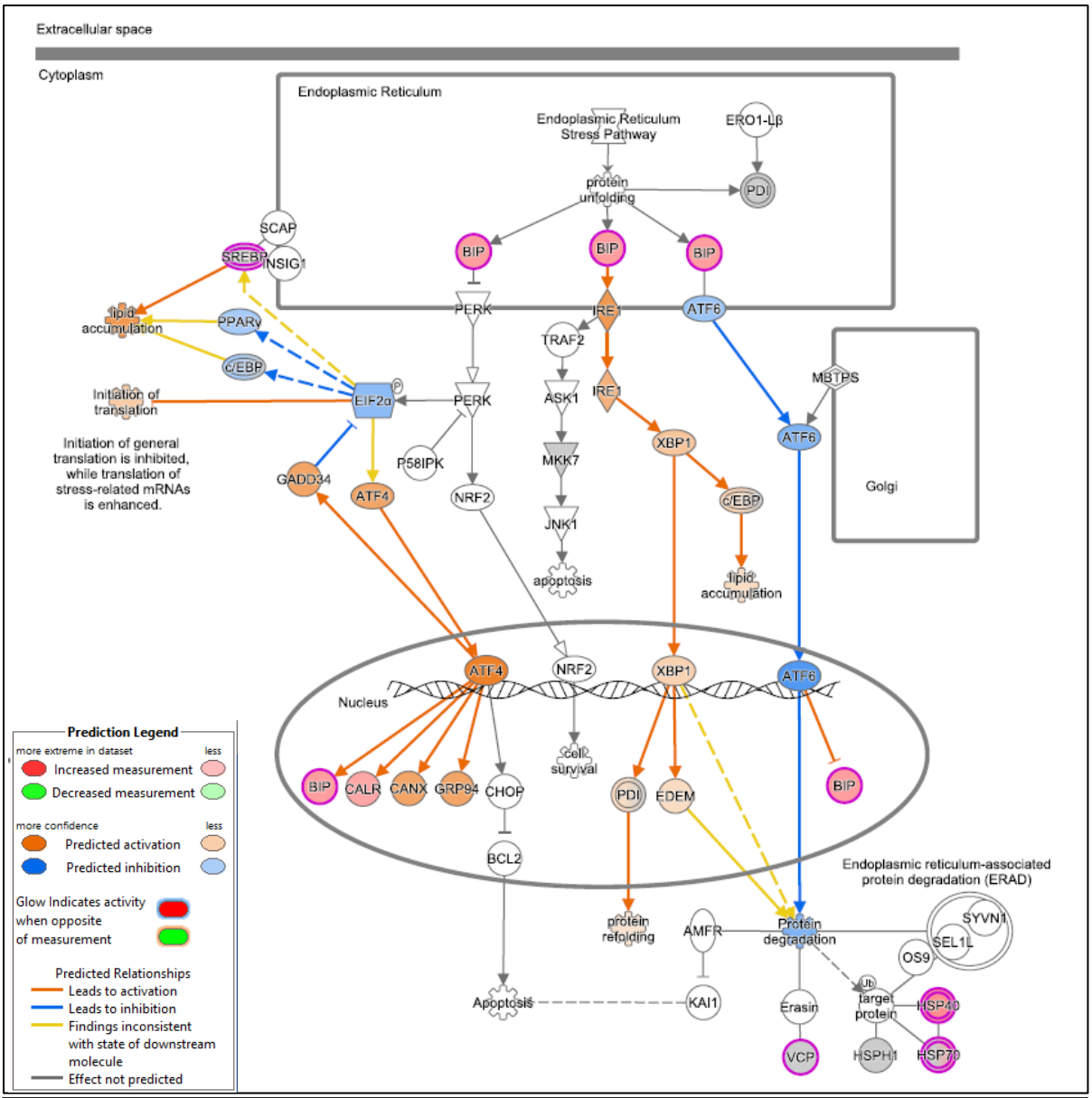

Figure 5. 8 Schematic drawings of the enrichment of SAFB1 and SAFB2 within the canonical Unfolded Protein Response pathway. Magenta lines represent proteins with which SAFB1 and SAFB2 are interacting, according to our results.

5.2.1.3.3 SAFB1 and SAFB2 in Huntington's Disease Signalling

SAFB proteins are highly expressed in the brain [356], especially in the hippocampus [30], and they are abnormally expressed in Huntington's Disease [95]. Huntington's Disease Signalling is also a significantly enriched pathway for both SAFB1 and SAFB2 interacting proteins. The pathway presents a -log(p-value)=1.63 (significantly enriched) and a zScore=2.0 for SAFB1 and a -log(p-value)=1.48 (significantly enriched) and no zScore for SAFB2.

The co-IP results revealed the following proteins as SAFB1 interactors: ATP5F1C (p=0.026, FDR=0.05), ATP5PB (p=0.043, FDR=0.065), CLTC (p=0.012, FDR=0.037), DCTN1 (p=0.018, FDR=0.043), EGFR (p=0.009, FDR=0.035), HDAC1 (p=0.026, FDR=0.05), HDAC2 (p=0.043, FDR=0.065), HSP70 family protein 1B, BIP, HSP70 family protein 8 and 9, POLR2A (p=0.011, FDR=0.037), POLR2B (p=0.024, FDR=0.048), PRKCI (p=0.023, FDR=0.047), PRKCZ (p=0.009, FDR=0.035), TBP (p=0.003, FDR=0.031) and UBA52 (p=0.02 , FDR=0.045), as shown in Erro! Fonte de referência não encontrada..

On the other hand, SAFB2 seems to be interacting to similar proteins as SAFB1, such as: ATP5F1C (p=0.017, FDR=0.061), ATP5F1E (p=0.017, FDR=0.061), ATP5PB (p=0.033, FDR=0.072), CLTC (p=0.016, FDR=0.061), DCTN1 (p=0.012, FDR=0.06), EGFR (p=0.013, FDR=0.06), HDAC1 (p=0.013, FDR=0.06), HDAC2 (p=0.034 , FDR=0.073), HSP70 family protein 1B, BIP, HSP70 family protein 8, HSP70 family protein 9, POLR2A (p=0.02, FDR=0.061), POLR2B (p=0.013, FDR=0.06), as shown in Erro! Fonte de referência não encontrada..

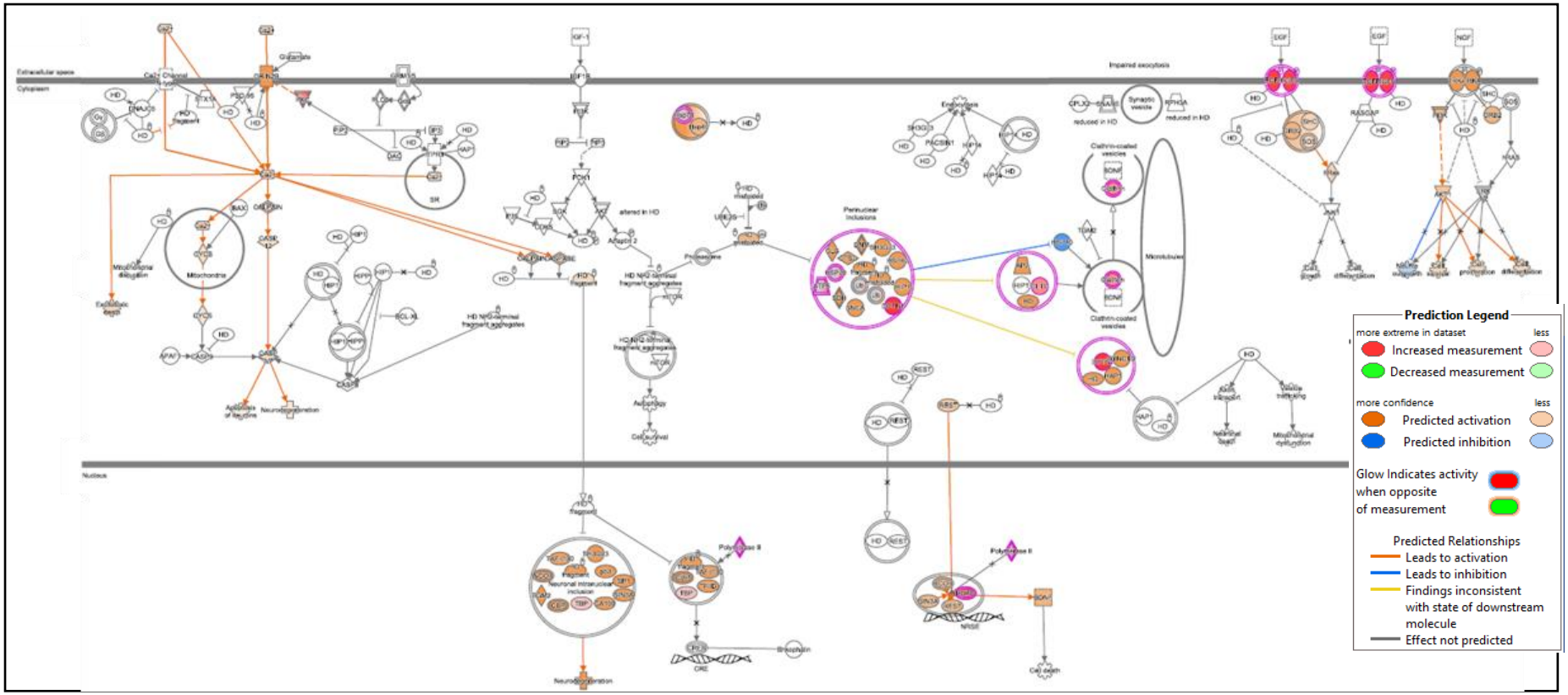

Figure 5. 9 SAFB1 enrichment within the Huntington's Disease Signalling Magenta lines represent a group of proteins with which SAFB1 is interacting according to our results

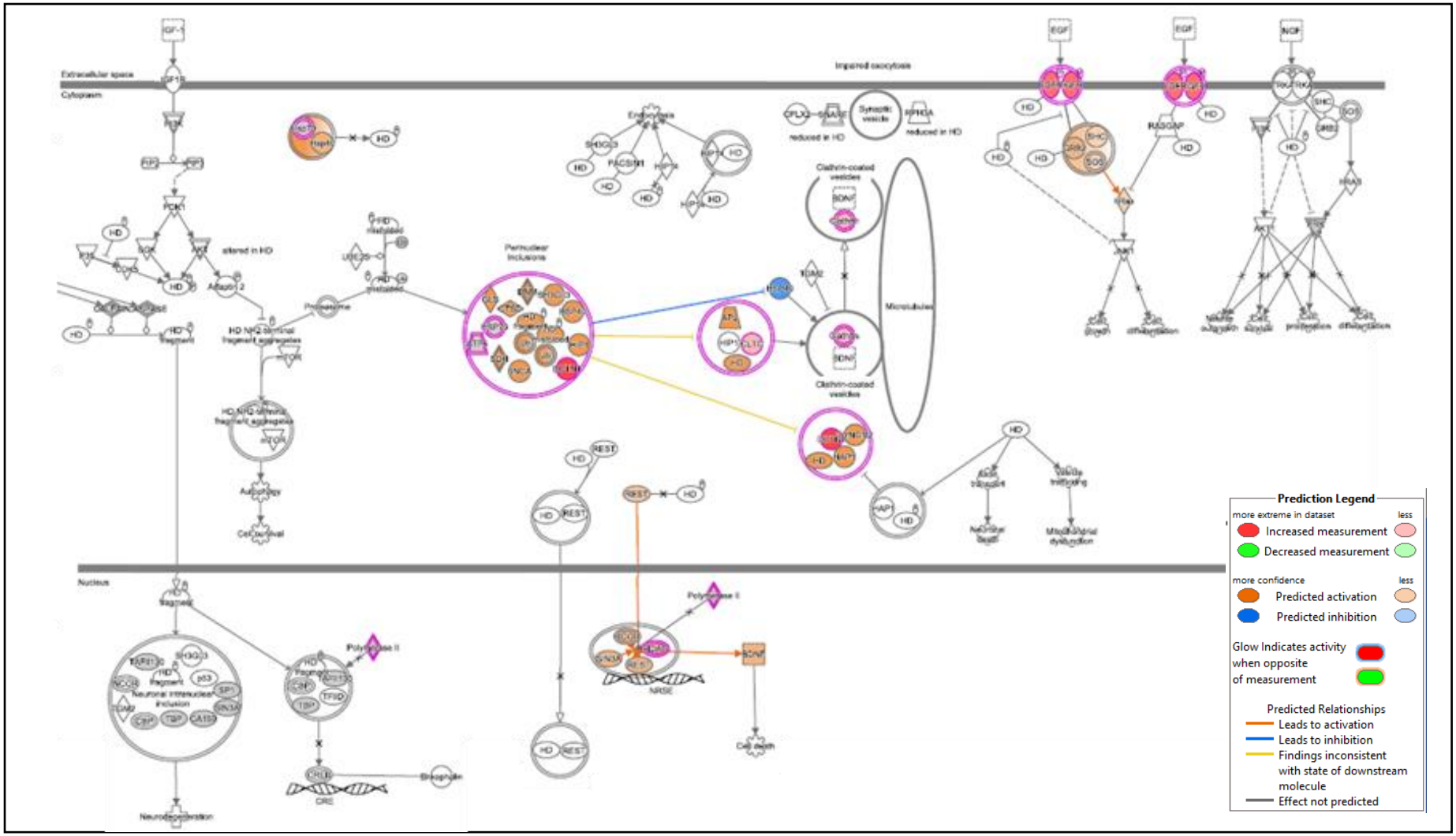

Figure 5. 10 SAFB2 enrichment within the Huntington's Disease Signalling. Magenta lines represent a group of proteins with which SAFB2 is interacting, according to our results.

5.2.1.4 SAFB1/SAFB2 Network analyses in HeLa cells

The IPA network analyses of SAFB1 and SAFB2 have shown an extremely high enrichment of both proteins within the RNA metabolism cluster. From the top five networks where SAFB1 was identified, all of them are involved in RNA post-transcriptional modifications, with scores>100. Other clusters identified were (i) protein synthesis, (ii) RNA damage and repair, (iii) gene expression, (iv) molecular transport/RNA trafficking, and (v) cancer/gastrointestinal diseases (Erro! Fonte de referência não encontrada.).

SAFB2 was shown to be involved in a more diverse range of clusters. Besides the RNA post-transcriptional modifications, which was also present in all categories, SAFB2 is not only associated with the same clusters as SAFB1 (protein synthesis, RNA damage and repair, gene expression, molecular transport/RNA trafficking) but in (i) cancer/cell death and survival, (ii) cellular assembling and organisation/function and maintenance, (iii) DNA replication/recombination and (iv) cell cycle, all of them with scores>100 (Erro! Fonte de referência não encontrada.).

Top networks	Score	Focus Molecules	Molecules
RNA Post-Transcriptional Modification, Protein Synthesis, RNA Damage and Repair	200	136	ABT1, AHCTF1, ANKRD28, ASCC3, BBX, BMS1, BYSL, CACTIN, CHMP1B, CUL1, CXorf56, DCAF13, DDX10, DDX18, DDX21, DDX47, DDX52, DDX54, DGCR8, DHX37, DHX57, DKC1, DNTTIP2, EFTUD2, EIF2AK2, EIF2S1, EIF3, EIF4B, EIF6, ESF1, EXOSC10, FAM98A, FCF1, FRG1, G3BP2, GAR1, GEMIN5, GLYR1, GNL3, H1-10, H1-2, H1-4, H3C1, HEXIM1, HNRNPA3, HNRNPR, HP1BP3, IFI16, IGF2BP3, ILF2,ILF3, IMP4, KRI1, LAS1L, MCM5, MPHOSPH10, MYBBP1A, NAT10, NCAPH, NFkB (complex), NGDN, NHP2, NKRF, NOB1, NOC4L, NOL7, NOL9, NOP10, NOP14, NPM1, NSD2, NUP153, NVL, PABPC4, PDCD11, PHF6, PINX1, PNO1, PWP1, RBM4, RCL1, RECQL4, RPF1, RPP40, RPS11, RPS13, RPS15, RPS15A, RPS16, RPS17, RPS18, RPS19, RPS2, RPS20, RPS23, RPS24, RPS25, RPS26, RPS27, RPS27L, RPS28, RPS3, RPS3A, RPS4X, RPS5, RPS9, RPSA, RRP12, RRP1B, RSL1D1, Ribosomal 40s subunit, Rnr, SART3, SEC13, SENP3, SIRT7, SPTY2D1, SRFBP1, STK10, SUCLG1, TARDBP, TBL3, TEX10, TFG, TMPO, TOP1, TRMT1L, UBTF, UTP11, UTP14A, UTP15, UTP3, WDR3, WDR36, WDR74, WDR75, XRCC6, XRN2, ZFR, ZNF668.
Protein Synthesis, RNA Damage and Repair, RNA Post-Transcriptional Modification	197	135	60S ribosomal subunit, AGK, ARMCX3, AURKAIP1, BOP1, BRIX1, CGAS, CNBP, DAP3, DDX24, DDX27, DDX56, DHX30, DIMT1, Dgk, EBNA1BP2, EIF3E, ERAL1, ERK1/2, Eif2, FBL, FXR1, FXR2, GNL2, GPATCH4, GPATCH8, GRSF1, GTPBP4, HNRNPU, INF2, KRR1, LARP7, LYAR, MRM1, MRM3, MRPL10, MRPL11, MRPL14, MRPL15, MRPL16, MRPL17, MRPL20, MRPL21, MRPL22, MRPL27, MRPL3, MRPL30, MRPL32, MRPL38, MRPL4, MRPL41, MRPL47, MRPL48, MRPL58, MRPL9, MRPS11, MRPS15, MRPS16, MRPS17, MRPS18A, MRPS18C, MRPS2, MRPS22, MRPS23, MRPS24, MRPS26, MRPS27, MRPS31, MRPS34, MRPS35, MRPS5, MRPS6, MRPS7, MRPS9, NIFK, NOP53, NOP56, PCBP1, PDGF (family), PES1, POP1, PPAN, PRR11, PTCD3, RBM28, RC3H1, RPF2, RPL10, RPL10A, RPL12, RPL13, RPL13A, RPL14, RPL15, RPL18, RPL18A, RPL21, RPL22, RPL23, RPL23A, RPL24, RPL26, RPL26L1, RPL27, RPL27A, RPL28, RPL29, RPL3, RPL30, RPL31, RPL32, RPL35A, RPL36, RPL36A, RPL37, RPL38, RPL4, RPL6, RPL7, RPL7A, RPL8, RPL9, RPLP0, RPS6, RPS8, RRS1, SERBP1, SRP68, SRP72, STAU1, STAU2, STRBP, TRA2A, TRIP4, TSR3, UBA52, WDR46, YTHDC2, ZC3HAV1, ZCCHC3.
RNA Post-Transcriptional Modification, Protein Synthesis, Gene Expression	186	131	48s, AGO2, AKAP, AKAP17A, AKAP9, ANKRD17, ARGONAUTE, ASH2L, ATXN2, ATXN2L, CAPRIN1, CCAR2, CELF1, CMSS1, COPA, CPSF1, CPSF4, CPSF6, CPSF7, CSDE1, CSTF3, Cpsf, DAZAP1, DDX1, DDX17, DDX20, DDX3X, DDX5, DDX6, DHX9, DIDO1, DROSHA, EDC3, EGFR, EIF3B, EIF3C, EIF3D, EIF3G, EIF3H, EIF4A, EIF4E, EIF4F, EIF4G1, ERI1, EWSR1, Eif4g, FIP1L1, FMR1, FUS, GPATCH2, HDAC2, HELZ, HERC2, HNRNPD, HNRNPDL, HNRNPF, HNRNPH1, HNRNPLL, HNRNPM, IGF2BP1, IGF2BP2, ITCH, KIFC1, LARP4, LARP4B, MBNL1, MCM3, MIB1, MKRN1, NEURL4, NFX1, NONO, NOP58, NTHL1, NTPCR, NUDT21, NUFIP2, OTUD4, P4HA1, PABPC1, PABPN1, PCBP2, PELO, PKP2, POLR2A, PPP1R10, PRRC2A, PRRC2B, PRRC2C, PTBP1, PUF60, PUM1, PURA, RBBP6, RBM10, RBM14, RBMS1, RIOK2, RPA2, RPS6KA, RPUSD3, RTCB, RTRAF, SEC16A, SF1, SFPQ, SLBP, SMG6, SMN1/SMN2, SPATS2, SSBP1, SYNCRIP, TARBP2, TCOF1, TDRD3, TFE3, TIAL1, TNRC6A, TOP3B, TOR4A, TRMT2A, Topoisomerase, U2SURP, UBAP2L, UPF1, UPF2, WDR33, WWP1, WWP2, XPO1, XRN1, YBX1, YTHDF1, YTHDF2, YTHDF3, ZC3H4, ZC3H7A, ZC3H7B, ZNF121, ZNF598.
RNA Post-Transcriptional Modification, Molecular Transport, RNA Trafficking	181	129	ACIN1, ALYREF, AQR, Akt, BCLAF1, BTBD10, CASC3, CCDC9, CCNT1, CDC40, CDC5L, CDK4/6, CDK5RAP1, CFAP20, CHD3, CHD4, CHERP, CHTOP, CLK1, CLK3, CRNKL1, DDX23, DDX39B, DDX46, DHX35, DHX8, DNA-PK, EIF4A3, FAM120A, FUBP3, HNRNPA0, HNRNPA1, HNRNPC, HNRNPH2, HNRNPH3, HNRNPUL1, HNRNPUL2, KIF18A, LARP1, LSM12, LSM14A, LSM14B, LUC7L, LUC7L2, LUC7L3, MATR3, MKRN2, NCBP1, NKAPD1, NKTR, NOL6, P-TEFb, PAXBP1, PNN,P OLDIP3, PPHLN1, PPIG, PPIH, PRPF19, PRPF3, PRPF31, PRPF38B, PRPF40A, PRPF4B, PRPF6, PRPF8,PSPC1, PYHIN1, RALY, RBM12B, RBM15, RBM15B, RBM22, RBM25, RBM27, RBM39, RBM6, RBM8A, RBMX, RBMXL1, RNPS1, RRP7A, RSRC1, S6K1, SAFB, SAFB2, SART1, SF3A2, SF3B2, SF3B3, SLTM, SNRNP200, SNRNP40, SNRNP70, SNRPA, SNRPA1, SNRPB, SNRPC,SNRPD1, SNRPD2, SNRPD3, SNW1, SREK1, SRPK1, SRPK2, SRRM1, SRRM2, SRRT, SRSF10, SRSF3,SRSF6, SRSF7, SRSF8, SRSF9, STRAP, SUZ12, TAF15, TASOR, THOC1, THRAP3, TTC3, U1 snRNP, U2 snRNP, U2AF1/U2AF1L5, U2AF2, U4 snRNP, U5 snRNP, U6 snRNP, UPF3B, WDR83, WTAP, XAB2, YLPM1, YTHDC1, ZC3H13, ZC3H14, ZC3H18, ZFC3H1, ZNF326, snRNP.
RNA Post-Transcriptional Modification, Cancer, Gastrointestinal Disease	166	123	ABCE1, ABCF1, ABCF2, ARL6IP4, ASPH, ATP synthase, ATP5, ATP5F1A, ATP5F1C, ATP5ME, ATP5PB, ATXN10, B3GALT6, BUD31, CCAR1, CCNL1, CEP78, CKAP4, CWC25, CWF19L2, DDX41, DHX15, DPM1, EGLN,E LMSAN1, EPB41L4B, EPB41L5, ESCO2, F1 ATPase, FAM111B, FAM133B, FARS2, GOLGA2, GPATCH1, GRAMD1B, GRAMD2B, H2BC13, H2BU1, HMGN2, IFRD1, IMP3, ITPR, JPH1, KIAA1522, KIF11, KIFC3, LACTB, LGALS3BP, LIG3, LMNA, LRRC59, MAGED2, MAGOHB, MAIP1, MBD4, MID1, MKI67, MORF4L2, MOV10, MTDH, MYCBP2, Mitochondrial complex 1,Mpt pore, NCAPD3, NCOA5, NDUFA4, NFIC, NHS, NUFIP1, NUMA1, NXF1, OSBPL3, Oxphos, PARP, PARP1, PARP12, PEG10, PFKM, PFKP, PHC2, PIMREG, PLEC, POGZ, PPP2R2A, PSMD3, RAB11FIP5, RAB14, RACK1, RAE1, RBMS2, RICTOR, RRBP1, SCAF11, SF3B1, SF3B6, SLC25A24, SLC25A5, SLC25A6, SLU7, SMC2, SPECC1L, SPOUT1, STRN3, TBC1D10B, TCAF1, TCF25, TECR, TFAM, TFIP11, THYN1, TIMM44, TJP1, TNKS, TRIM27, TRIM56, TUFM, TWNK, U2af, UIMC1, UQCRC2, USP7, VIRMA,Vegf, WDR70, XRCC1, XRCC1-LIG3-POLbeta-PARP1, XRCC1-LIG3-POLbeta-PARP1-PNKP, YBX3,ZBTB11, ZFP2, ZGPAT, ZMAT2, ZNF221, ZNF281, ZNF460, ZNHIT6, adenosine-tetraphosphatase, cytochrome C, cytochrome-c oxidase, pentosyltransferase.

Figure 5. 11 Top 5 High-scoring networks (Score >100) enriched for SAFB1 identified by Ingenuity Pathway Analysis in HeLa cells out of 11 pathways significantly enriched.

Top networks	Score	Focus Molecules	Molecules
Protein Synthesis, RNA Damage and Repair, RNA Post-Transcriptional Modification	187	127	48s,60S ribosomal subunit, ALYREF, CAPRIN1, CCDC9, CCNL1, CHD3, CHD4, CNBP, DAP3, DDX17, DDX39B, DDX3X, DDX50, DDX6, DHX15, DHX30, EIF3, EIF3B, EIF3D, EIF3G, EIF3H, EIF4A, EIF4A1, EIF4A3, EIF4B, EIF4E, EIF4F, EIF4G1, ERH, ERI1, ERK1/2, Eif4g, FAM120A, FBL, FUBP3, GRSF1, HEXIM1, HNRNPA0, HNRNPA1, HNRNPA3, HNRNPDL, HNRNPR, HNRNPU, HNRNPUL2, IGF2BP2, IGF2BP3, ILF2, LARP1, LARP4B, LARP7, MAGOHB, MATR3, MKRN1, MRPL58, MRPS11, MRPS2, MRPS24, MRPS25, MRPS27, MRPS31, MRPS34, MRPS35, MRPS5, MRPS7, NCBP1, NKRF, PABPC1, PABPC4, PABPN1, PCBP1, PCBP2, PDGF (family), PKC alpha/beta, POLDIP3, PTBP1, PTCD3, PYHIN1, RALY, RBM8A, RBMX, RC3H1, RNPS1, RPL10A, RPL13, RPL15, RPL18, RPL18A, RPL21, RPL24, RPL26, RPL26L1, RPL27, RPL3, RPL31, RPL35A, RPL36, RPL36A, RPL4, RPL6, RPL7, RPL7A, RPL8, RPL9, RPS6KA, RPS6KA4, Ribosomal 40s subunit, S6K1, SAFB, SF3A2, SLBP, SMG6, SNRNP70, SRPK2, SRRT, SRSF3, SRSF7, SRSF9, STAU1, STAU2, STRBP, TAF15, TARDBP, TDRD3, THRAP3, TRA2A, U2AF2, UPF1, UPF3A, UPF3B, YBX1, YTHDC1, ZC3H18, ZC3H7A, ZC3HAV1, ZCCHC3, ZFC3H1, ZFR, ZNF326, hnRNP H.
RNA Post-Transcriptional Modification, Cancer, Cell Death and Survival	185	126	ABCF2, ACOT9, ACTB, APOB, AQR, ARGONAUTE, ASCC2, ASPH, ASPM, BBX, BUD31, C1orf174, CEP78, CLK1, CLK3, Calcineurin A, Calmodulin, DEK, DRG1, EEF1AKNMT, ELAVL1, EPRS1, EXOSC10, EXOSC9, FAF2, FAM111B, FSCN1, FTSJ1, GOLGA2, GTF3C1, GTF3C3, H2BC13, HNRNPA2B1, IGF2BP1, KARS1, KIF11, KIR2DL2, KPNA2, LSM12, LSM14A, LUC7L3, LUZP1, MBD4, MORF4L2, MOV10, MTCH1, MTHFD2, MTREX, MYH9, NEB, NFX1, NKTR, NUDT16L1, NUFIP1, P4HA1, PA2G4, PAXBP1, PDCD11, PHC2, PNN, PPIG, PPIL2, PRKAA1, PRPF31, PRPF38B, PRPF4B, PRPF6, PRSS23, PSMD3, RAB14, RARS1, RBM15, RBM22, RBM33, RBM6, RBMS2, RCN2, RICTOR, RPL12, RPL13A, RPL22, RPL23, RPL28, RPL37, RPL38, RSL24D1, SART1, SEC61A1, SEPTIN7, SF3B1, SF3B2, SHMT2, SMOC1, SNRNP200, SNRNP40, SNRPA, SNRPA1, SNRPB, SNRPD1, SNRPD2, SNRPD3, SON, SRBD1, SRP14, SSB, STRN3, Serine Protease, TBC1D10B, TCF12, TFIP11, THAP11, THYN1, TMEM33, TRAF4, TRIM56, TRIP6, TRMT10C, TRMT1L, TRMT2A, TWF1, U1 snRNP, U2af, U4 snRNP, U5 snRNP, U6 snRNP, UIMC1, Ubiquitin, VCP, WBP11, XAB2, YBX3, ZC3H13, ZFP2, ZGPAT, ZNF460, ZNF629, alcohol group acceptor phosphotransferase, methyltransferase, peptidylprolyl isomerase, snRNP.
Protein Synthesis, RNA Post-Transcriptional Modification, Gene Expression	165	118	ACADM, AGO2, AHNAK2, ATP synthase, ATP5, ATP5F1A, ATP5F1C, ATP5F1E, ATP5PB, ATXN2, Ant, BYSL, CCAR2, CD3 group, COPA, CPSF1, CPSF4, CPSF6, CPSF7, CSDE1, CSTF3, Coup-Tf, Cpsf, Creb, DDX1, DIDO1, EIF1AX, EIF2B4, EIF2S1, EIF2S2, EIF2S3, EIF5B, ERAL1, ERK, Eif2, F1 ATPase, FAM98A, FMR1, FXR2, H1-0, HERC2, HSPA8, IK, IMP3, ITPR, KIFC1, KIFC3, KLHL12, LACTB, LSG1, LSM14B, Laminin1, MAIP1, MRPL10, MRPL11, MRPL15, MRPL16, MRPL21, MRPL22, MRPL27, MRPL3, MRPL30, MRPL32, MRPL38, MRPL4, MRPL41, MRPL47, MRPL48, MRPS18A, MSI2, Mitochondrial complex 1, Mpt pore, NCOA5, NDUFA4, NEURL4, NHS, NOB1, NONO, NUDT21, OTUD4, Oxphos, PEPCK, PFKM, PFKP, PITX1, PNO1, PPP1R10, PRRC2A, PUM1, PURA, RBFOX1, RBM17, RBM26, RBM27, RBMS1, RPRD2, RTCB, RTRAF, SAFB2, SCAF11, SLC16A3, SLC25A11, SLC25A13, SLC25A3, SLC25A5, SLC25A6, SPATA5, SPATA5L1, SPATS2, SPECC1L, SPOUT1, SSBP1, SSR1, STRAP, SYNCRIP, T3-TR-RXR, TCAF1, TECR, TFAM, TMA16, TNRC6A, TOMM20, TUFM, TWNK, Transportin, U2SURP, UBAP2L, UQCRC2, UTP14A, VIRMA, WDR33, XPO1, YTHDF1, YTHDF3, ZC3H4, ZNF121, adenosine-tetraphosphatase, cytochrome C, cytochrome-c oxidase,eIF.
RNA Post-Transcriptional Modification, Cellular Assembly and Organization, Cellular Function and Maintenance	158	115	14-3-3, APC (complex), ARHGAP21, ARHGEF2, ATAD3A, AURKB, Akt, Alpha actin, Alpha tubulin, Ap2 alpha, BETA TUBULIN, BUB3, CAMSAP2, CAPZB, CCT3, CCT4, CCT5, CDC20, CDK13, CDK4/6, CENPE, CEP170, CEP55, CKAP2, CLASP2, CNP, CUL1, Cytoplasmic Dynein, DAZAP1, DCTN1, DDX5, DLGAP5, DNA-PK, DNAJ, DNAJA, DNAJA1, DNAJA2, DNAJA3, DNAJB11, DNAJC21, DNAJC9, DROSHA, DYNC1H1, DYNLL1, Dynein, E2f, EFTUD2, EMD, FAM83D, FXR1, GPATCH1, Gamma tubulin, HCFC1, HNRNPD, HNRNPF, HNRNPH1, HNRNPH2, HNRNPH3, HNRNPM, HNRNPUL1, HSP, HSPA1A/HSPA1B, HSPA9, Hdac, Hsp22/Hsp40/Hsp90, Hsp70, Hsp90, ILF3, INCENP, ITIH2, KIF1C, KIF23, KIF2A, KNSTRN, KSR1, LGALS3BP, MAGED2, MAP7, MAP7D1, MAP7D3, MARK1, MARK2, MARK3, MIB1, MICU1, MTCL1, MYL6, NACA, NAP1L1, NQO1, NUDC, NUMA1, PARD3, PCNT, PLK1, PPHLN1, PPP2R2A, PRC1, PRPF19, PRPF8, PRPS1, PSPC1, Plk, RACGAP1, RAE1, RBM10, RBM4, RRP36, RSRC1, SF1, SF3B3, SLC25A1, SNRPC, SNW1, SRRM2, SRSF8, STT3B, SUZ12, TCP1, TIMM50, TTC3, TUBA1C, TUBA4A, TUBB, TUBB3, TUBB4B, TUBG1, U2AF1/U2AF1L5, UTP18, Vdac, WDR5, WDR82, WDR83, YAP/TAZ, YLPM1, YWHAG, YWHAZ, ZC2HC1A, tubulin, tubulin (family).
RNA Post-Transcriptional Modification, RNA Damage and Repair, Protein Synthesis	142	108	ADRB, AHNAK, AKAP, AKAP13, AKAP17A, AKAP9, ANKRD17,AP-3, AP2A1, AP2B1, AP2M1, AP3D1, AP3M1, APC/APC2, ARF3, ARF4, ASCC3, ATXN2L, AURKA, Adaptor protein, Adaptor protein 2, Ap2, Arf, BMS1, BRIX1, Beta Arrestin, CCDC86, CFAP20, CK1, CLTC, COP I ,COPB1, CPS1, CSNK1A1, CSNK1E, CTPS1, Cd1, Ck2, Clathrin, DAPK3, DDX10, DDX18, DDX21, DDX24, DDX27, DDX52, DDX54, DGCR8, DHX57, DKC1, DNTTIP1, DNTTIP2, Dishevelled, Dynamin, EBNA1BP2, EIF3E, FBXO11, FLNB, G3BP2, GLYR1, GNL2, GNL3L, GPATCH4, GTPBP4, H3C1, HDAC2, HMMR, HP1BP3, Histone H1, IFI16, IMP4, INF2, Importin alpha, Importin alpha/beta, Importin beta, KPNA4,KRR1, LYAR, MAP1LC3, MKI67, MPHOSPH10, MRTO4, MYBBP1A, MYO1E, Mrlc, NAT10, NFkB (complex), NIFK, NOP53, NUFIP2, NUP153, Nucleoporin, PFN1, PINX1, POM121/POM121C, POP1, PP1 protein complex group, PP2A, PRKAR2A, PRKRA, PRRC2C, PWP1, Pkar2, Pld, RNMT, RPF2, RPL10, RPL14, RPS10, RPS12, RPS16, RPS19, RPS2, RPS20, RPS27, RPS3A, RPS4X, RPS6, RPS9, RPSA, RRP12, RRP1B, RRS1, Rnr, SART3, SERBP1, SIRT7, SPATS2L, SPTY2D1, SRP68, SRP72, Smad, Sos, TARBP2, TASOR, TPX2, WDR46, YTHDC2, ZNF622, atypical protein kinase C.

DNA Replication, Recombination, and Repair, RNA Post-Transcriptional Modification, Cell Cycle	132	103	ABCD3, ABCE1, ABCF1, AHCTF1, ATP5PO, ATPase, BCLAF1, Basc, CCNT1, CDC6, CDK12, CELF1, CHERP, CHTOP, CMSS1, Cbp/p300, DDX20, DHX8, DHX9, DNA Polymerase, DNA polymerase delta, DNA-directed RNA polymerase, DNA2, ECT2, EGFR, EIF6, ELOA, ERCC3, ERCC5, ESCO2, EWSR1, FEN1, GTF2H1, GTF2H3, H1-10, H2AZ1, HBS1L, HDAC1, HISTONE, HLTF, HNRNPC, Hat, Hdac1/2, Histone h3, Histone h4, Holo RNA polymerase II, KAT7, KIF14, KIF1B, Keratin II, 6, LIG3, LONP1, LRPPRC, LRRC59, MACF1, MARS1, MCM3, MCM5, MCM6, MCM7, MORF4L1, MPG, MRE11, MRN, MRPS17, MSH6, MTA2, Mcm, Mi2, Mta, MutS alpha,NCAPD3,NCAPH,NOL7,NTHL1,NuRD,P glycoprotein,P-TEFb, PARP1, PCNA, PDHA1, PELO, PHF6, POLR1B, POLR1C, POLR2A, POLR2B, POLRMT, POLbeta-POLepsilon-POLgamma-XRCC1-LIGI-PARP1-PCNA-FEN1, PRKDC, Postincision complex 3, RAD50, RAI14, RBM15B, RBMXL1, RCC2, RCL1, REEP4, RFC1, RFC3, RIOK2, RNA polymerase I, RNA polymerase II, RPA, RSL1D1, RUVBL1, Rfc, SFPQ, SMARCA5, SMC2,S MC4, SMN1/SMN2, SND1, SSRP1, SUPT16H, Sin3, TCOF1, TFIIH, THOC1, TIP60, TNKS, TOP1,TOP2A, TOP3B, Top2, Topoisomerase, UBTF, UHRF1, UTP15, UTP4, WDR75, XRCC1, XRCC1-LIG3, XRCC1-LIG3-POLbeta-PARP1, XRCC1-LIG3-POLbeta-PARP1-PNKP, XRN2, YY1, ZC3H11A, histone deacetylase.

Figure 5. 12 Top 6 High-scoring networks (Score >100) enriched for SAFB2 identified by Ingenuity Pathway Analysis in HeLa cells out of 10 pathways significantly enriched.

5.2.1.5 SAFB1/SAFB2 Enrichment in Bio functions and Disease

The IPA analyses for Bio functions and Disease found SAFB1/SAFB2 enrichment in 254 and 121 bioprocesses, respectively, excluding cancer-related functions. Most of these bio functions involved RNA processing, corroborating with the network analyses shown previously. In addition, SAFB1 and SAFB2 presented similar associations within the categories and functions, with SAFB1 having slightly lower p values and a higher number of associated proteins than SAFB2. Top functions where RNA/mRNA processing, splicing and alternative mRNA splicing are crucial for protein expression and protein variability; rRNA processing and modifications, protein translation; protein metabolism (translation, synthesis, elongation, expression and termination); and molecular trafficking. **Erro! Fonte de referência não encontrada.** and **Erro! Fonte de referência não encontrada.** show the top categories, functions, p values and number of proteins with which SAFB1/2 were found to be associating with.

Categories	Diseases or Functions Annotation	p-Value	Activation z-score	# Molecules
RNA Post-Transcriptional Modification	Processing of RNA	8.66E-183	0.146	244
	Splicing of RNA	1.18E-113	-1.556	154
	Processing of mRNA	1.97E-105	-2.115	150
	Splicing of mRNA	1.43E-101	-2.162	135
	Modification of rRNA	1.41E-66	2.764	71
	Processing of rRNA	4.69E-65	2.764	69
	Alternative splicing of mRNA	7.09E-22		29
RNA Damage and Repair	Decay of mRNA	4.87E-72	2.415	76
	Nonsense-mediated mRNA decay	7.83E-72	1.969	74
	Metabolism of RNA	3.58E-19		32
Protein Synthesis	Translation	6.08E-109	-0.198	176
	Synthesis of protein	2.04E-98	3.317	203
	Expression of protein	7.40E-98	-0.713	179
	Initiation of translation of protein	6.69E-67	-1.816	80
	Metabolism of protein	1.05E-62	3.918	255
	Termination of translation of protein	1.58E-34		45
	Elongation of protein	9.50E-31		47
Molecular Transport, RNA Trafficking	Transport of RNA	4.06E-31	2.749	47
	Transport of mRNA	1.47E-30	2.756	45
	Export of mRNA	3.05E-30	2.752	43
	Nuclear export of mRNA	6.77E-30	1.964	42
Molecular Transport	Nuclear export	1.12E-29	2.2	49
	Nuclear export of molecule	6.53E-27	1.964	45
	Export of molecule	1.40E-11	2.95	54
Gene Expression, Protein Synthesis	Translation of RNA	1.83E-31	-4.615	63
Gene Expression	Expression of mRNA	1.28E-27	-3.495	71
	Expression of RNA	2.64E-24	0.816	293
	Transcription	2.43E-15	2.633	251
	Transcription of RNA	1.04E-14	1.522	218
	Transcription of DNA	1.47E-12	0.053	182
Gastrointestinal Disease, Organismal Injury and Abnormalities	Colorectal disorder	1.05E-12	-1.629	469
	Liver lesion	6.22E-14	0.135	472
DNA Replication, Recombination, and Repair	Repair of DNA	3.35E-14	3.396	64
Developmental Disorder, Embryonic Development, Organismal Survival	Death of embryo	6.72E-13	-5.791	34
Cell Death and Survival	Necrosis	1.22E-12	-5.365	287
Cell Cycle, DNA Replication, Recombination, and Repair	DNA recombination	2.48E-13	3.698	39
	Homologous recombination of cells	1.16E-10	4.679	24
	Recombination of cells	2.00E-10	4.679	25
Cell Cycle, Cellular Assembly and Organization, DNA Replication, Recombination, and Repair	Segregation of chromosomes	1.66E-14	1.871	34
Cell Cycle	Cell cycle progression	2.16E-11	2.427	133

Figure 5. 13 Top 40 Diseases and Bio functions enriched for SAFB1.

Categories	Diseases or Functions Annotation	p-Value	Activation z-score	# Mole
RNA Post-Transcriptional Modification	Processing of RNA	1.81E-129	-0.158	187
	Splicing of RNA	1.28E-91	-1.137	127
	Processing of mRNA	3.90E-83	-1.678	122
	Splicing of mRNA	3.52E-80	-1.904	110
	Modification of rRNA	1.28E-34		45
	Processing of rRNA	3.51E-34		44
	Alternative splicing of mRNA	3.60E-19		25
RNA Damage and Repair	Nonsense-mediated mRNA decay	7.64E-47		54
	Decay of mRNA	2.07E-46	1.969	55
	Metabolism of RNA	1.86E-18		29
Protein Synthesis	Translation	9.26E-83	-0.603	140
	Synthesis of protein	5.50E-78	2.687	165
	Expression of protein	5.03E-74	-1.027	142
	Initiation of translation of protein	4.02E-51	-1.816	64
	Metabolism of protein	3.38E-50	3.213	208
	Termination of translation of protein	2.72E-24		34
	Elongation of protein	3.52E-22		36
Organismal Survival	Organismal death	1.65E-12	-12.989	227
Molecular Transport, RNA Trafficking	Transport of RNA	2.09E-21	2.592	35
	Export of mRNA	8.22E-21	2.63	32
	Nuclear export of RNA	8.22E-21		32
	Transport of mRNA	1.66E-20	2.621	33
	Nuclear export of mRNA	2.64E-20		31
Molecular Transport	Nuclear export	6.25E-20		36
	Nuclear export of molecule	4.72E-18		33
Gene Expression, Protein Synthesis	Translation of RNA	5.70E-25	-4.289	51
	Translation of mRNA	1.23E-24	-4.289	50
	Initiation of translation of mRNA	1.34E-12	-1.993	18
Gene Expression	Expression of mRNA	2.31E-22	-3.379	58
	Expression of RNA	8.41E-22	0.888	246
	Transcription	5.97E-13	2.74	207
	Transcription of RNA	1.77E-12	1.58	180
Gastrointestinal Disease, Organismal Injury and Abnormalities	Colorectal disorder	1.02E-13	-1.97	400
	Liver lesion	1.10E-12	0.093	393
DNA Replication, Recombination, and Repair	Repair of DNA	4.66E-17	3.909	62
Developmental Disorder, Embryonic Development, Organismal Survival	Death of embryo	3.75E-16	-5.871	35
Cellular Assembly and Organization, Cellular Function and Maintenance	Organization of mitotic spindle	1.94E-13		21
Cell Death and Survival	Necrosis	1.40E-14	-6.023	252
	Cell death of tumor cell lines	3.09E-14	-4.297	171
Cell Cycle, DNA Replication, Recombination, and Repair	DNA recombination	4.16E-13	3.049	35

Figure 5. 14 Top 40 Diseases and Bio functions enriched for SAFB2

5.2.2 TMT-MS Validation – Ago2 Pulldown

An interesting finding from the TMT-MS is the interaction between SAFB proteins and argonaute 2, the RISC catalytic component (Ago2, gene ID: 27161). To confirm SAFB1 and SAFB2 bound Ago2 proteins, further immunoprecipitation experiments were carried out using a commercially available mouse monoclonal Ago2 antibody (#018-22021, Fujifilm, USA) and recently developed custom made anti-SAFB1 and anti-SAFB2 antibodies (detailed in methods, section 2.20).

Firstly, the SAFB1 and SAFB2 custom made antibodies were validated by western blotting. Then, both antibodies were raised in Rabbit against the following unique peptides:

SAFB1: aa 478 to 491(SVEKAKNEPVGKKT) and aa 609 to 626 (KEPRKSRDSESHSRVRER).

SAFB2: aa 907 to 953 (AQGGHSQGHVVPGGGLEGGGVASQDRGSRVPHPHPHPPPYPHFTY).

Erro! Fonte de referência não encontrada. shows the western blot results from SH-SY5Y cells treated with siRNA for SAFB1, SAFB2, SAFB1/SAFB2 knockdown and control, probed with the custom made anti-SAFB1 (Ab1895) and anti SAFB2 (Ab1898) antibodies (methods, section 2.7 – 2.8). Single SAFB1 knockdown reduced SAFB1's expression to 34.2%, whereas in the KD SAFB1, expression was reduced to 48.0% compared to the non-targeting control (A). Similarly, SAFB2 expression was reduced to 56.6% and 49.6% in single and double KDs.

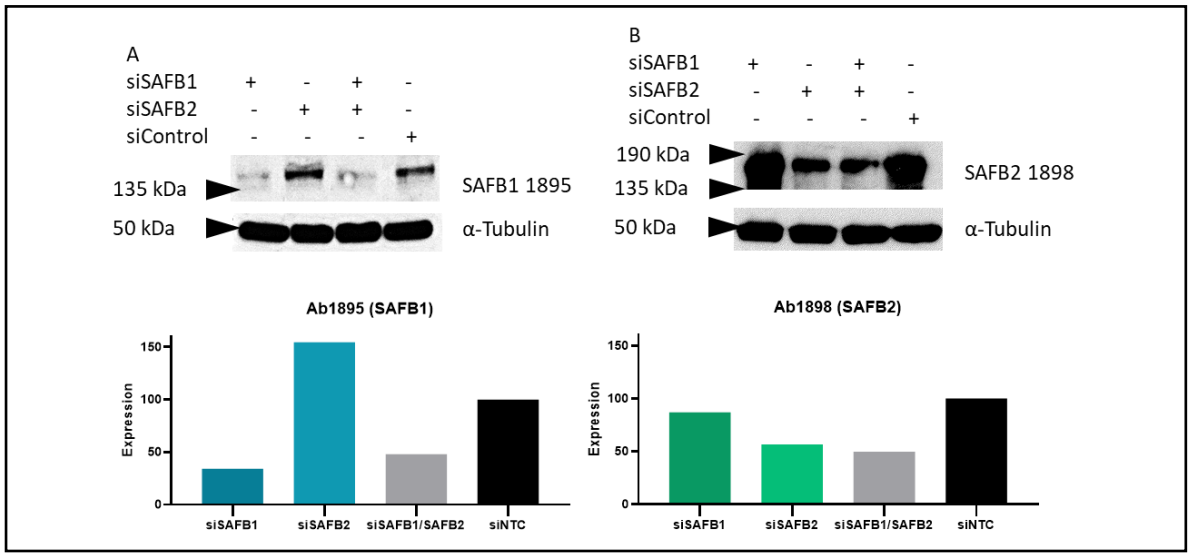

Figure 5. 15 Validation of custom-made Rabbit anti-SAFB1 and anti-SAFB2 antibodies. SH-SY5Y cells were treated with 60 pMol of siRNA (siSAFB1, siSAFB2, siSAFB1+SAFB2 and siNTC) 48 hrs before harvest. (A) Western blot results for SAFB1 1595 antibody (top) and protein levels (bottom); (B) Western blot results for SAFB2 1598 antibody (top) and protein levels (bottom).

5.2.2.1 Co-Immunoprecipitation Ago2, SAFB2 and SAFB2

We sought to confirm the presence of Ago2 protein among the protein complexes co-immunoprecipitated with SAFB1 and SAFB2 antibodies and vice-versa. In addition, we included a heat stress condition to observe any changes in SAFB1/2 interaction with Ago2 upon the stress. Co-immunoprecipitation was carried out in SH-SY5Y cells, submitted or not to 60 minutes of heat stress to observe possible changes between Ago2 and SAFB1 or SAFB2 during heat stress (methods section 2.4, 2.8). All antibodies were cross-linked to the magnetic beads to minimise the unspecific background due to the heavy (50 kDa) and light (25 kDa) antibody chains. Details on the protocol are described in methods section 2.17. Three independent IP assays were carried out, and western blots were probed with either Ago2, SAFB1 or SAFB2 antibodies. Western blot samples are presented as 5.5% of input, 50% of the pulldown fraction (PD) and 5.5% of the immune-depleted fraction (ID) in control and heat shock samples. The samples were either normalised to the input densities when bands were detectable or corrected to the gel density. Paired T-tests and 2-way ANOVAs were used to assess statistical significance. Moreover, the co-IP for SAFB1/2 and Ago2 detection was performed on a different cell line from the one line used in the proteomic analysis. While the TMT-MS were performed in a HeLa cell line, the SAFB1/2 and Ago2 co-IP were performed in the SH-SY5Y cells, reinforcing the robustness of the analysis and suggesting that the expression and interaction between SAFB1/2 and Ago2 is not cell type-specific.

The results presented in **Erro! Fonte de referência não encontrada.** show that the IP was successful for the control samples (Ago2 IP/Ago2 probe, SAFB1 IP/SAFB1 probe and SAFB2 IP SAFB2 probe). The Ago2 pulldown (**Erro! Fonte de referência não encontrada.** A) shows two detectable bands at ~110 kDa and 65 kDa. Ago2 has two isoforms with similar molecular weights (97 and 93.4 kDa). However, a 65 kDa band is observed in the heat-shocked and control samples, suggesting cleavage of the protein. Moreover, this low MW band (65 kDa) and the higher MW (uncleaved) Ago2 bands are more prevalent following HS than in NHS conditions (p=0.001) shown in **Erro! Fonte de referência não encontrada.** A (red boxes) and B. No Ago2 bands could be detected in the immunodepleted lane, suggesting that the beads predominantly captured ago2.

SAFB1 pulldown also shows two bands at 190 and 110 kDa (**Erro! Fonte de referência não encontrada.** C) before and after heat stress induction, suggesting constitutive cleavage of the protein. No significant changes can be seen in the densities between NHS and HSNR samples (**Erro! Fonte de referência não encontrada.** D).

Similarly, SAFB2 pulldown also shows two bands at ~190 and ~110 kDa (**Erro! Fonte de referência não encontrada.** E) before and after heat stress induction, although the lightweight bands are almost not detected (red boxes). No significant changes can be seen in the densities between NHS and HSNR samples (**Erro! Fonte de referência não encontrada.** F).

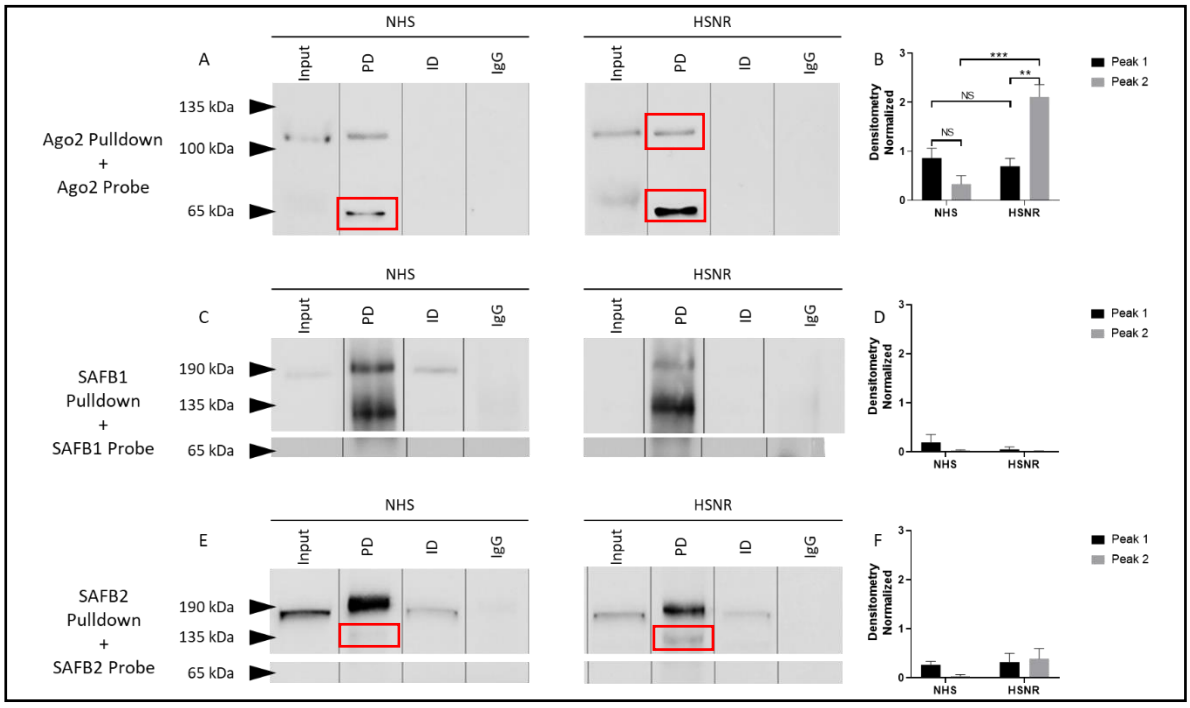

Figure 5. 16 IP was performed with 1200 µg of pre-cleared SH-SY5Y cells lysate with either (A) Ago2 (Fujifilm, Japan), (C) SAFB1 (Invitrogen), (E) SAFB2 (Invitrogen), or IgG (Invitrogen) antibodies cross-linked magnetic beads. The immunoprecipitate was run on Western blots and probed with Ago2 (Fuji Film, Japan), SAFB1 (Bethyl) and SAFB2 (Bethyl) antibodies. Red boxes highlight the main differences in bands densities. Input and immunodepleted (ID) lysates were run as controls. (B, D and F) Densitometry quantitation of detectable bands. Co-IP was performed on three independent experiments. Only one representative blot is shown per IP.

When the SH-SY5Y cell lysates are immunoprecipitated with Ago2 antibody and probed with SAFB1, the western blot results show that SAFB1 is co-immunoprecipitated Ago2 in the absence of stress. SAFB1 was also found in the Ago2 immunodepleted lane, indicating that SAFB1 was not solely bound to Ago2. Interestingly, the IP band has a molecular weight of ~110 kDa. However, this interaction is not seen when samples are submitted to heat stress (**Erro! Fonte de referência não encontrada.** A). No significant changes can be seen in the densities between NHS and HSNR samples (**Erro! Fonte de referência não encontrada.** B).

Meanwhile, SH-SY5Y cell lysates immunoprecipitated with Ago2 antibody and probed with SAFB2 show only a faint band at no bands at ~200 kDa in the NHS samples, whereas IP bands in the HSNR samples are barely detectable (red boxes), Suggesting that SAFB2 has minimum interaction with Ago2. The presence of strong SAFB2 bands in the immunodepleted lanes reinforces the evidence of minimal interaction between Ago2-SAFB2. (**Erro! Fonte de referência não encontrada.** C). No significant changes can be seen in the densities between NHS and HSNR samples (**Erro! Fonte de referência não encontrada.** D).

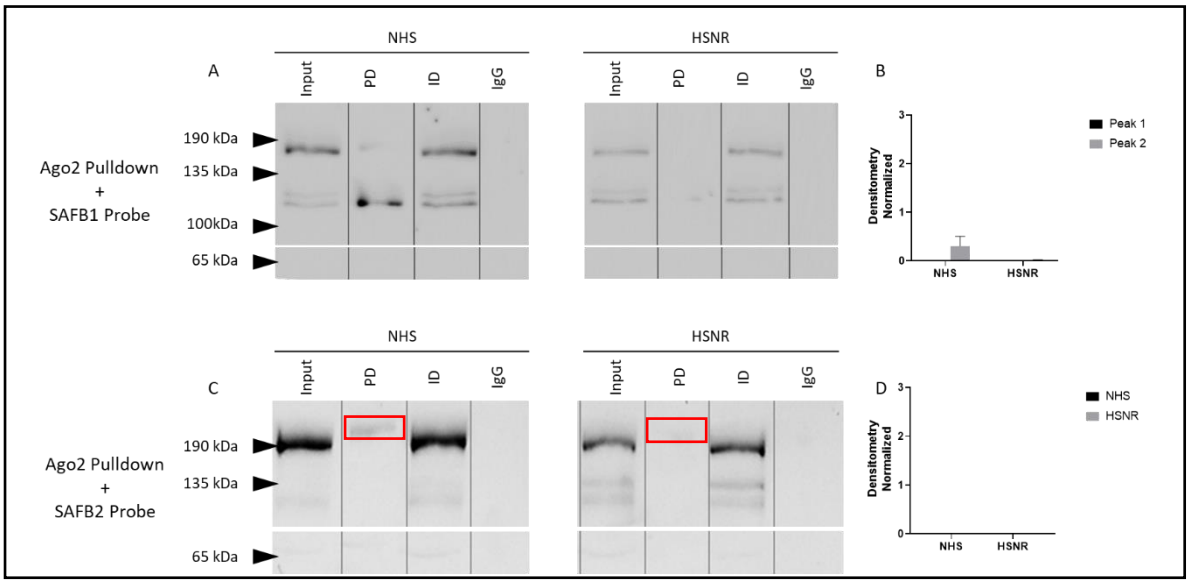

Figure 5. 17 IP was performed with 920 µg of pre-cleared SH-SY5Y cells lysate with Ago2 (Fujifilm, Japan) and IgG (Invitrogen) antibodies cross-linked magnetic beads. The immunoprecipitate was run on Western blots and probed with (A) SAFB1 (Bethyl) or (C) SAFB2 (Bethyl) antibodies. Red boxes highlight weak bands densities. Input and immunodepleted (ID) lysates were run as controls. (B and D) Densitometry quantitation of detectable bands. Co-IP was performed on three independent experiments. Only one representative blot is shown per IP.

On the other hand, when the cell lysates are immunoprecipitated with SAFB1 antibody and probed with Ago2, the western blot results show that Ago2 is strongly co-immunoprecipitated SAFB1 in the absence of stress but weakly present following heat stress. Similar to the previous results, the Ago2 band is at ~110 kDa. However, the absence of bands at the input and immunodepleted lanes suggests low amounts of Ago2 interacting with SAFB1 (**Erro! Fonte de referência não encontrada.** A). Moreover, there is a significant difference in the densities observed in NHS and HSNR bands ($p=0.0275$, **Erro! Fonte de referência não encontrada.** B).

Pulldown of SAFB2 protein followed by Ago2 gel probing shows co-immunoprecipitation of Ago2 with SAFB2 as well, with bands at ~110 kDa. Once again, Ago2 is strongly co-immunoprecipitated with SAF21 in the absence of stress but is barely present in heat stress samples. The absence of bands at the input and immunodepleted lanes suggests low amounts of Ago2 interacting with SAFB2 (**Erro! Fonte de referência não encontrada.**

C). Moreover, no significant changes can be seen in the densities between NHS and HSNR samples (Erro! Fonte de referência não encontrada. D).

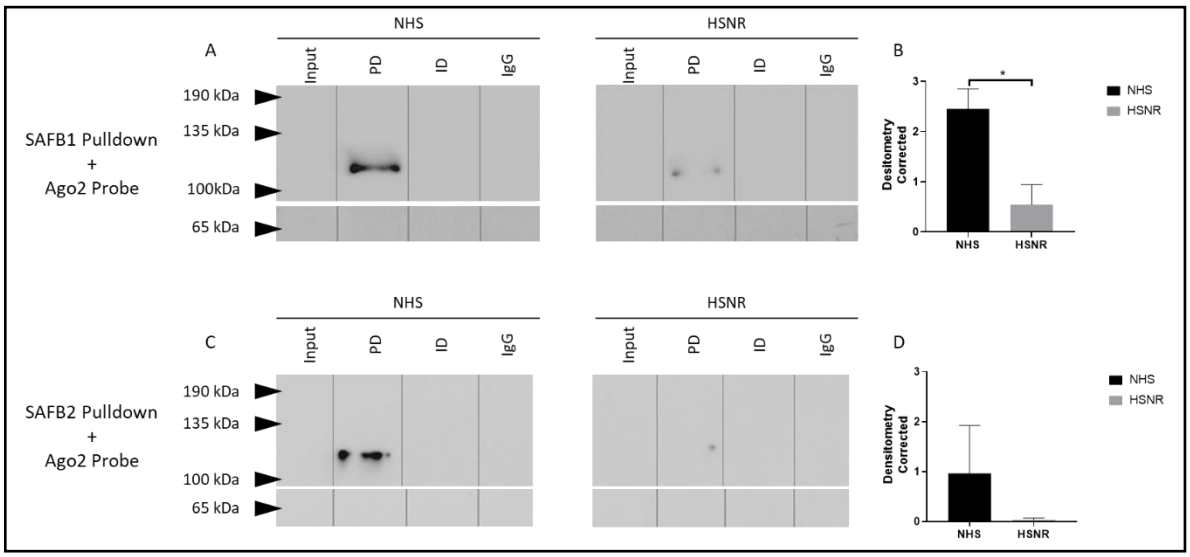

Figure 5. 18 IP was performed with 1200 µg of pre-cleared SH-SY5Y cells lysate with either (A) SAFB1 or (C) SAFB2 (Invitrogen), and IgG (Invitrogen) antibodies cross-linked magnetic beads. The immunoprecipitate was run on Western blots and probed with (A) SAFB1 (Bethyl) or (C) SAFB2 (Bethyl) antibodies. Input and immunodepleted (ID) lysates were run as controls. (B and D) Densitometry quantitation of detectable bands. Co-IP was performed on three independent experiments. Only one representative blot is shown per IP.

5.3 Summary

Proteomic analyses by TMT is a powerful means of identifying novel protein interactions and involvement in biological processes and diseases. For example, using immunoprecipitation, mass spectrometry and proteomics analyses, novel SAFB1 and SAFB2 binding partners were found with functions related to RNA processing, previous corroborating studies showing the involvement of SAFB proteins in miRNA metabolism and splicing [49, 127, 332]. Moreover, SAFB1 and SAFB2 proteins interact with important proteins in pathways not described in the literature before, such as eIF2 signalling, the unfolded protein response and Huntington Disease (HD) signalling.

SAFB1 was already shown to be involved in several processes, such as chromatin organization [9, 127], transcription [14, 79, 127], stress response [1, 15], DNA

damage/repair [11], apoptosis [43, 82] and diseases [88, 95, 212, 357]. SAFB1 and SAFB2 share the same protein domains with an overall similarity of 74% between their sequences, which could explain both overlapping functions and the 784 binding partners found in common for both proteins in the results here presented. However, this similarity imposed a challenge to identify unique functions for proteins in the past, as old antibodies could not distinguish between SAFB1 and SAFB2. Using custom made antibodies capable of differentiating between the two proteins, we have found that SAFB2 interacts with 94 binding partners who are not interacting with SAFB1. Among these are four subunits of the T-complex protein 1 (TCP) ring complex (TRiC), a multi-protein complex chaperone that folds polypeptides [324, 325], being responsible for ~10% of the proteome folding [326, 327].

Moreover, the subunit epsilon (CCT5), also found to interact exclusively with SAFB2, was shown to inhibit the aggregation of mutant huntingtin proteins (mHTT) in Huntington disease [358], linking SAFB2 to the HD. Finally, the tubulin-alpha4A (TUBA4A) is another unique SAFB2 interactor that is a highly conserved homologue of a rat testis-specific protein. Moreover, mutations in the TUBA4A gene have been linked to cases of familial ALS [328]. Interestingly, SAFB2 -/- KO mice showed increased testis weight associated with an increased number of Sertoli cells, which is not seen on SAFB1-/- KO mice, reinforcing a unique role SAFB2 in the differentiation and activity of Sertoli cells [31]. Hence, our results showed that SAFB2, but not SAFB1, is significantly enriched among the canonical pathway of the oestrogen Receptor Signalling, probably acting as an activator of the process.

Previous studies have already shown the involvement of SAFB1/2 in the regulation of transcription. SAFB1 binds and represses HSP27 promoter activity [7], the androgen receptor (AR) [32] and p53 [50]. Moreover, SAFB1 and SAFB2 knockdown lead to the downregulation of 259 genes and upregulation of another 457 genes in cancer cells [43], whereas decreased SAFB1 expression resulted in down-regulation of the skeletal muscle genes myogenin and Brg1 [45] as well as XOR [79]. Notably, the TMT findings showed that SAFB1 and SAFB2 interact with Ago2, which was confirmed by further co-immunoprecipitation experiments. Ago2 is crucial for all miRNA mediated translational silencing, and several accessory proteins modulate its function. Ago2 is mainly a

cytoplasmatic protein, although a few studies have shown that it can be found in the nucleic compartment in smaller amounts [359, 360], and it co-immunoprecipitates with euchromatin in senescent cells [361]. Ago2 have been shown to relocate to stress granules and P bodies in the cytoplasm upon stress conditions, to which the heat-shock protein 90 (HSP90) activity was required for efficient recruitment [362]. The protein activator of the double-stranded RNA activated protein kinase (PRKRA or PACT) IS a Dicer cofactor that plays a vital role in eIF2 signalling/ISR, is also recruited to the stress granules [362]. Here we have shown that SAFB2 is a binding partner of PACT, whereas SAFB1 binds PRK, a binding partner of PACT that initiates the eIF2 signalling/ISR upon dsRNA presence during viral infections [363]. Whether the interaction between SAFB1/2 and Ago2 occurs within the nucleus or cytoplasm remains unknown.

SAFB1 and SAFB2 were also found to be interacting partners of several important regulators of miRNA transcription: Drosha, DGCR8, Ago2, Exportin-1 and TNRC6A. Thus, these data further support a role for SAFB1 and SAFB2 in miRNA processing.

Previous studies have placed SAFB1 as part of spliceosomes [73], which are snRNA/snRNP complexes that aids introns extrusion [364], therefore implicating SAFB1 in splicing. Our laboratory has previously used individual-nucleotide resolution cross-linking and immunoprecipitation (iCLIP) analyses to identify SAFB protein-RNA interactions [30], supporting the involvement of SAFB1 in the regulation of processing and splicing. This work showed that GAAGA was the most significantly enriched pentamer sequence bound by SAFB1 and that the core AGA, GAA, or AAG motifs in exons were bound to regulate splicing. In addition, the knockdown of either SAFB1 and SAFB2 significantly reduced the processing of miR-19A from the miR-92 cluster (miR-17,18a,19a,20a,19b,92) and that the simultaneous knockdown of SAFB1 and SAFB2 further decreased miR19A processing. We have also knocked down SAFB1 and SAFB2 in SH-SY5Y neuroblastoma cells and completed transcriptomic analyses measuring the expression levels of small RNAs (including miRNAs) and mRNAs. These analyses found that SAFB1/2 proteins regulated the expression of specific miRNAs (e.g. let-7A1, let-7A3 and miR-98). The proteomic analyses described in this chapter found SAFB proteins interacted with Drosha, DGCR8, Ago2, MOV10 and FMR1. Subsequent confirmatory pulldown assays with Ago2 and blotting with SAFB1 and SAFB2 specific Abs confirmed

that both SAFB2 and SAFB1 interacted with Ago2. Together with our previous iCLIP results (showing SAFB1 binds the miR-17-92 cluster), these results suggest SAFB proteins may bind miRNAs and/or regulate processing by microprocessor and Ago2. SAFB2 and SAFB1 were recently shown to bind microprocessors to mediate the processing of miRNA sub-optimal stem-loop structures [332, 365]. A further study has shown that SAFB proteins interact with LC3 and are loaded into extracellular vesicles along with their RNA and protein cargos [366]. Together these results suggest a significant function of the SAFB family of proteins is to regulate the processing and function of miRNAs. Therefore, it is crucial to carry out further experiments to investigate the role SAFB1 and SAFB2 play in regulating microprocessors, miRNA export and processing by the Dicer complex; Ago2 mediated repression of translation.

In addition, there is considerable evidence linking the aetiology of polyQ disorders to the altered function of miRNAs, Ago2 and Pum1 [96-101]. The expression of ATXN1 and ATXN3 was regulated by specific miRNAs whose altered expression increased disease pathology [102, 103]. The RBP (and Ago2 binding protein) PUM1 was recently shown to regulate ATXN1 expression directly, and its downregulation increased neurodegeneration [99]. Expression of mutant HTT also leads to a Dicer-mediated increase in the levels of small CAG-repeated RNAs (sCAGs) that caused neuronal death via an Ago2 dependent mechanism [98]. Furthermore, HTT has also been shown to be involved in RNA transport in neuronal dendrites where it associates with Ago2 to repress translation [96, 97]. Importantly, aggregation of mutant HTT (mHTT) was shown to impair LC3 mediated autophagy leading to Ago2 accumulation and global alterations in miRNA levels and activity [100].

Our analyses of constitutively bound proteins confirmed known SAFB1/2 protein partners and identified novel partners involved in miRNA biogenesis (e.g., Ago2, DROSHA, MOV10, FMR1) and implicated in the aetiology of polyglutamine diseases (e.g., HDAC1,2, HSP90, HSC70, HSP40s, Pum1). Both SAFB1 and SAFB2 interacted with DGCR8, Drosha and Ago2 and pulldown assays with Ago2 and blotting with SAFB 1 and SAFB2 specific Abs confirmed SAFB2 and SAFB1 interacted with Ago2. Furthermore, our group has shown pulldown assays, and immunocytochemical analyses using a SAFB1/2 antibody showed SAFB proteins were associated with the Dicer-TRBP complex

and that SAFB1/2 are found co-localised with Dicer in the cytoplasm. SAFB proteins were also found to bind specific miRNAs (e.g. Let-7, miR-199, miR-23) [367, 368], and we found the overexpression of miR-23 mediated a functional increase between SAFB1/2 and the Dicer/TRBP complex (communication, Dr Youn-Bok Lee). Furthermore, SAFB2 and SAFB1 were recently shown to bind microprocessors to mediate miRNA sub-optimal stem-loop structures [51, 365]. A further study has shown that SAFB proteins interact with LC3 and are loaded into extracellular vesicles along with their RNA and protein cargos[366]. These data suggest that SAFB proteins regulate miRNA processing, Ago2 function and vesicular transport, and their abnormal expression will contribute to SCA and HD pathology.

To investigate Ago2 function, luciferase reporter assays for miRNAs identified as regulated by SAFB1/2 could be used and activity measured following SAFB1/2 knockdown. In addition, interactions with MOV10 and FMR1 could also be investigated using co-immunoprecipitation.

eIF2 signalling

Under normal conditions, the eIF2 complex recognises the AUG codon in mRNA and facilitates the translation initiation in the ribosomes. However, phosphorylation of serine 51 at the eIF2α subunit by one of the four eIF2α kinases occurs in response to various stresses and reduces AUG recognition downregulating the synthesis of new AUG initiated proteins during the stress and activating ATF4 [353]. Interestingly, heat shock proteins such as HSF1 and HSPs are not AUG initiated, enabling them to be synthesised and upregulated during the stress [369].

Moreover, our results show that SAFB1 binds either directly or indirectly to the double-stranded RNA-dependent protein kinase (EIF2AK2 or PRK), one of the four eIF2α kinases responsible for the initiation of the cascade. Once PKR is activated in the presence of double stranded-RNA, it phosphorylates eIF2α at Serine 51, triggering the pathway leading to protein downregulation [363, 370, 371]. On the other hand, SAFB2 was shown to interact with PRKRA (PACT), which in mouse activates PKR in the absence of dsRNA, triggering the pathway [372, 373]. Interestingly, the interaction of Ago2 and PRKRA has been already shown in the literature [374].

The Unfolded Protein Response

Proteins are synthesised by the ribosomes and mainly secreted to the lumen of the Endoplasmic Reticulum (ER), where a network of chaperones, glycosylating enzymes, and oxido-reductases guide the 3D shape folding of these newly synthesised proteins [375]. Following synthesis and folding, proteins are directed to their final destinations. However, this process is not simple and involves energy consumption to run correctly. Therefore, cells must keep track of the quality of the proteins made. Unfolded and misfolded proteins can occur due to various reasons, but the accumulation of them triggers the activation of stress responses and can lead to apoptosis if not controlled [376]. Hence, cells have developed distinct ways to sense the accumulation of mis/unfolded proteins. The PERK, IREα and ATF6 are the three unfolded protein sensors within the ER that trigger the UPR [377]. PERK leads to the activation of eIF2α and activates ATF4, linking the UPR to the ISR [378]. Activation of the UPR triggers two temporally distinct cellular events to mitigate protein misfolding: an initial reaction to reduce protein synthesis and enhance degradation of misfolded proteins and the second wave of transcriptional upregulation of hundreds of target genes involved in global proteostasis control [354]. In mammalian cells, PERK and IRE1a are kept inactive in a complex ER chaperone BiP/Grp78 in an inactive monomeric state [379]. BiP binding to mis/unfolded proteins releases PERK and IRE1a. Notably, the accumulation of abnormal protein aggregates comprises the first steps of several neurodegenerative diseases, such as Alzheimer, Parkinson, Huntington's diseases and amyotrophic lateral sclerosis (ALS) [380]. Our findings show that SAFB1, SAFB2 and BiP are interacting partners in homeostasis, placing both SAFB1 and SAFB proteins at the beginning of the UPR. Thus, one can hypothesise that upon stress, SAFB1/2 are recruited to the nSBs in the nucleus, and SAFB1/2 dissociation from the BiP/PERK/IRE1a can somehow alter its stability, leading to the release of BiP, PERK and IRE1α and initiating the UPR. However, more experiments are needed in order to validate this hypothesis.

Huntington Disease signalling

Another recently published study from our laboratory shows SAFB1 is linked to Huntington's Disease (HD) and spinocerebellar ataxias (SCA) [95]. HD is caused by abnormal CAG repeat expansion in exon 1 of the Huntingtin gene (HTT), generating an expanded polyglutamine (polyQ) tract, which becomes pathological when it exceeds 35 repeats [381]. The resulting mutated protein (mHTT) is cleaved into smaller toxic fragments that aggregate and accumulates in Striatal GABAergic spiny projection neurons (SNPs) in the striatum, disrupting their biological functions and leading to neuronal damage [382]. The longer the repeats, the earlier is the age of onset [383]. Spinocerebellar ataxia 1 (SCA1) is another polyQ disease in which the CAG repeat expands within the Ataxin-1 (ATXN1) gene. Greater than 39 uninterrupted CAG repeats within the ATXN1 gene are considered pathological [384], leading to toxic protein accumulation in the nucleus and disrupting neuron homeostasis. Cerebellum and brain from SCA1 patients exhibit signs of degeneration, alongside the loss of Purkinje cells caused by the development of aberrant protein aggregates [385, 386].

The study from the author's laboratory found abnormal cytoplasmic SAFB1 staining, coupled with increased nuclear SAFB1 staining in the striatal neurons of HD patients, which was not present in control patients. A similar pattern, with increased nuclear SAFB1 staining and abnormal cytoplasmatic SAFB1 staining, was seen in Purkinje cells in SCA patients, suggesting an association between increased abnormal SAFB1 expression in the cytoplasm and striatal/Purkinje cell injury. Interestingly, the proteomics analyses found SAFB1 and SAFB2 enriched within Huntington Disease signalling. The co-IP experiments showed a subset of proteins involved in this pathway to interact with SAFB1 and SAFB2 in HeLa cells, such as BiP, HSP70, ATP5F1C, CLTC, DCTN1, EGFR, HDAC1/2, POLR2A/B, PRKCI/Z, TBP and UBA52, corroborating with the results presented by Buckner [95]. Moreover, the iCLIP experiments conducted by the same authors reviewed that SAFB1 binds ATXN1 RNA at CAG repeats and the ATXN1 anti-sense RNA (GTC repeat). It has been proposed that long glutamine RNA tracts act as a scaffold to recruit and sequester SAFB1 (and other RNA binding proteins), thus disrupting SAFB1 balance within the cell. Since SAFB1 has roles in the stress response [1, 3, 11], RNA/miRNA processing [16, 30, 48, 127] and chromatin stability [190, 387]. Overall, the loss of SAFB1 function due to its abnormal expression would alter transcription, RNA processing and increase cell stress, contributing to neuronal cell death.

Additionally, since SAFB1 and SAFB2 are also involved in stress-related mechanisms, such as the UPR and the ISR, SAFB1 disruption is likely associated with cell pathology, especially in non-dividing cells. As neurons are post-mitotic and do not divide, they must have efficient stress acknowledgement, damage control, and repair mechanisms. Correct function of the UPR and eIF2 signalling allows the maintenance of cell homeostasis. If these mechanisms fail or are overloaded, these pathways will lead to apoptosis and cell death.

Chapter 06: Discussion

The aim of experiments described in chapter 3 was to investigate the dynamic recruitment (and distribution) of SAFB proteins to nSBs. To achieve this, nSBs were identified by the presence of HSF1, lncSATIII and SRSF1 in cells exposed to a heat stress. Additionally, we tested whether knocking down SAFB proteins would alter the transcription levels of the lncSATIII.

Previous studies have assessed the binding of SAFB1 to nSBs [1, 6, 11, 50, 74, 191, 202]. However, most of these studies relied on EGFP tagged deletion mutants and an anti-SAFB1/2 antibody. In this study, the specificity of the three rabbit anti-SAFB1/2 and SLTM antibodies was confirmed, and experiments then carried out. GFP-tagged SAFB1, SAFB2 and SLTM were overexpressed in SH-SY5Y cells, and heat stress was induced by keeping the cells at 42°C for 1hr. The GFP-tagged proteins were then monitored for up to 3.5 hrs post-stress in a live system (IncuCyteZoom) under normal conditions. Our results suggest that all overexpressed SAFB family members formed aggregates immediately after HS induction (1hr HS). Exogenous SAFB1 aggregates persisted for up to 1.5 hrs after cessation of the heat stress, while exogenous SAFB2 aggregates remained for up to 2.5 hrs after the heat stimulus was withdrawn. Exogenous SLTM aggregates persisted for more than 3.5 hrs of recovery. Thus, we determined optimal timepoints for further analysis immediately after heat stress (heat shock no recovery - HSNR) and one hour of recovery at normal conditions following heat stress (heat shock 1hr recovery - HS1hr). However, further conclusions are limited by the equipment IncuCyteZoom used to acquire images of the different time points. The IncuCyteZoom has a maximum lens magnification of 20x, which is not enough to clearly visualise subnuclear compartments. Therefore, the IncuCyteZoom was used to determine the SAFB aggregation time course, and the Opera system was used to unbiasedly measure SAFB nSB colocalisation.

Turning now to the experimental evidence of the characterisation of SAFB antibodies used in this study. Two methods were used to confirm the specificity of the anti-SAFB family member antibodies. First, the alignment of the epitope sequences targeted by each of the antibodies was accessed using the Serial Cloner 2.6.1 software. This alignment analysis revealed that the maximum level of similarity was no longer than two sequential amino acids on the peptide sequences of SAFB proteins taken from European

Bioinformatics Institute (EMBL-EBI) databases. Second, antibody specificity was confirmed by SAFB1, SAFB2 and SLTM siRNA mediated knockdown and followed by protein detection using western blotting. The results of the analyses of the western blots showed that the anti-SAFB1 and anti-SAFB2 antibodies bound their target protein and did not cross-react. These results show that the antibodies differentiate between the SAFB family members. Therefore, it can be assumed that the use of these antibodies in experiments reflects the characteristics of the specific SAFB family member being studied.

Several reports have shown that HSF1, Serine/arginine-rich splicing factor 1 (SRSF1), and lncSATIII colocalise with nSBs [179, 180, 192, 200]. To characterise SAFB proteins in nuclear stress bodies, antibodies targeting HSF1 and SRSF1 were used as stress markers for the immuno-characterization of nSBs following heat shock in SH-SY5Y cells. Following HS at different set time points, cells were fixed and stained for HSF1, SRSF1. Our results confirmed that the formation of nSBs containing either HSF1 or SRSF1 occurred following heat stress induction and persisted for one h after the heat stimulus was removed. Whereas aggregation of HSF1 or SRSF1 was not observed before heat induction.

Furthermore, another nSB/stress marker used was the identification of SATIII long non-coding RNA. RNA-FISH using a biotinylated-RNA probe targeting anti-sense SATIII showed that lncSATIII transcripts were absent before HS, but staining was visible following HS and persisted for one hour after the stress was withdrawn and our results are in accordance with previous reports [78, 180]. Having confirmed the presence of well-established stress markers HSF1, SRSF1 and lncSATIII in nSBs, the use of this protocol was then used to investigate the distribution of individual SAFB proteins following HS.

Prior studies have shown that SAFB proteins colocalise with HSF1 [1], SRSF1, RNA Pol II [127], and SATIII transcripts [9]. However, those studies were carried out using anti-SAFB antibodies that potentially cross-reacted with other family members. Therefore, previous studies do not provide a clear picture of the distribution and recruitment of SAFB1, SAFB2 and SLTM proteins during control, stress and post-stress or recovery conditions.

In this investigation, we assessed the distribution and recruitment of individual SAFB proteins to stress bodies using antibodies confirmed to be specific for each SAFB family member (see section 3.2.2). Cells were fixed and stained with anti-HSF1 and anti-SAFB1, anti-SAFB2 or anti-SLTM antibodies following HS at defined time points (no recovery and 1hr recovery). Our results show that before HS, all SAFB members show a diffuse granular distribution throughout the nucleus. Whereas following HS and recovery, SAFB1 form aggregates that do not colocalise well with HFS1 in nSBs. This means that SAFB1 may be re-directed upon stress, but not to the HSF1-positive nSBs. This finding is not consistent with previous reports of SAFB co-localisation with HSF1-positive nSBs after HS [1, 127, 191, 202]. However, the inconsistency may be due to their use of a non-specific anti-SAFB1 antibody.

On the other hand, SAFB2 was recruited immediately after HS and shows greater co-localisation with HSF1-positive nSBs and remains until later recovery times. Likewise, there was a change in SLTM distribution following HS, a few aggregates (less than SAFB2) were formed, which colocalised with HSF1-positive nSBs. In addition, co-staining SRSF1 with either SAFB1 or SAFB2 showed that similarly to the experiments using HSF1, SAFB2 shows greater co-localisation with SRSF1 than SAFB1. Taken together, these results show specific patterns of distribution and recruitment for SAFB1, SAFB2 and SLTM, suggesting that they have distinct roles within the cell

Furthermore, to quantify the effects of HS observed on the recruitment of SAFB proteins to $HSF1^{+}$ nSBs we developed an automated identification protocol using a Fiji/ImageJ MIA plugin that analysed a series of images and quantifies SAFB1, SAFB2 and SLTM in co-localisation with HSF1+ nSBs. Following previous studies [200], we defined two distinct populations: canonical $HSF1^{+}$ nSBs (large) and secondary $HSF1^{+}$ (small) foci by sorting their puncta depending on size. Our results show that in the absence of heat shock, HSF1 rarely forms aggregates and that following heat stress, with and without recovery, there was a significant rise in the formation of both small and large nSBs. Having confirmed the effects of heat stress induction, we moved on the quantify the distribution patterns of SAFB proteins. Our analysis showed that the distribution pattern of SAFB1 in the absence of stress consists of small puncta in ~40% of the cell population and large SAFB1 puncta in ~8% of the cell population.

Interestingly, no changes were observed in either large or small SAFB1 puncta formation immediately after HS. However, following 1hr of recovery, there was a statistically significant increase of large SAFB1$^+$ puncta in cells, but no significant changes in the proportion of small SAFB1$^+$ puncta. Moreover, in the absence of stress SAFB1 did not colocalise with HSF1$^+$ nSBs, whereas there was a small but statistically significant increase of co-localisation of SAFB1 with HSF1+ nSBs in 4.8% of the cell population immediately after HS and 5.7% following recovery. This finding is contrary to previous studies, which have suggested that SAFB proteins colocalise with nSBs in stressed cells. However, this can be explained by the difference in antibodies used. Previous studies [1, 6, 11, 74, 191, 202] used an antibody that recognised motifs shared between SAFB family members, whereas our results were obtained using antibodies specific for each family member.

In the case of SAFB2, we found large SAFB2$^+$ puncta in ~22% of cells and small SAFB2 puncta in ~81% of cells before stress induction. Following HS, we found no significant changes in the formation of large SAFB2$^+$ puncta immediately after HS, whereas the proportion of small SAFB2+ puncta decreased from ~81% to ~57.2%. Additionally, the results following one hour of recovery showed a dramatic increase in the formation of large SAFB2$^+$ puncta in ~51.7% of the cell population, whereas the percentage of small SAFB2$^+$ puncta returned to non-stress levels. Taken together, the results of our analysis suggest that SAFB1 and SAFB2 have different distribution patterns before and after HS, which provides evidence to support the idea that SAFB1 and SAFB2 have different roles in cell recovery from stress.

Regarding the recruitment of SAFB2 to HSF1$^+$ nSBs, the percentage of cells with large SAFB2+ puncta and HSF1 co-localisation increased to ~62% immediately after HS, further increasing to ~85.2% of cells after one hour of recovery. On the other hand, the small SAFB2+ puncta colocalised with HSF1+ nSBs showed an increase of only ~6% and ~8% of cells immediately after HS and following recovery, respectively. This suggests that the co-localisation between HSF1 and SAFB seen in old studies using the anti-HAP (SAFB1/2) antibody is probably due to SAFB2 recruitment to nSBs rather than SAFB1 and that SAFB2 might play a more critical role with the heat shock response than SAFB1. A possible explanation for this might be that SAFB2 interacts with HSF1 upon thermal

stress, and this interaction might be mediated via PPI domains (Gly-Rich and E/R-Rich) at the C-terminal end of SAFB2. This region is also where SAFB1 and SAFB2 diverge the most, which could explain why SAFB2, but not SAFB1, colocalization with HSF1+ nSBs is higher. In order to verify the hypothesis, proteins that interact with SAFB1 and SAFB2 were analysed (discussed below). This analysis did not find that HSF1 was among the binding partners of SAFB proteins, however, our analysis of SAFB interacting partners was performed under control conditions. Therefore, future work should examine the change in SAFB2 interacting partners to test whether the recruitment of SAFB2 to the HSF1+ nSBs is mediated by interactions between SAFB2 and HSF1.

Lastly, there were no significant changes in the percentage of cells exhibiting either large or small SLTM+ before, after HS, nor following one h of recovery. Moreover, the only significant change in co-localisation between HSF1+ nSBs cells and SLTM + puncta was an increase detected after one-hour recovery in large nSBs. Taken together, these results show that SAFB2 has greater recruitment to the canonical large HSF1+ nSBs following HS with and without compared to SAFB1 or SLTM, suggesting that SAFB2 has a more critical role in the HS recovery than SAFB1 or SLTM. As mentioned before, previous studies that had examined the co-localisation of SAFB with nSBs used a non-specific antibody [1, 6, 11, 74, 191, 202], making it impossible to differentiate the specific behaviour of SAFB family members following HS. Therefore, this study revealed that the SAFB co-localisation with HSF1+ nSBs following HS reported before is mainly related to the behaviour of SAFB2 rather than SAFB1 or SLTM.

Another significant aspect of nuclear body formation occurs by nucleating RNA and recruitment and accumulation of proteins [388]. Additional studies have also found that upon heat stress, HSF1 binds to the SATIII regions of the genome and recruits RNApolII [1, 74, 191], driving the transcription of this long non-coding RNA. These long transcripts remain and accumulate around the synthesis area, potentially attracting other RNA binding factors, such as SAFB2. This study found that lncSATIII transcripts form large and small aggregates in the nucleus following HS, with and without a recovery period, acting as the scaffold for the attachment of various proteins that form nSBs. These aggregates colocalise with HSF1, especially in the large nSBs population. On the other hand, our control experiments using a non-targeting RNA probe showed that despite the high debris

background content, this was more pronounced in the cytoplasm, an area that was excluded from the analysis. Hence, the formation of foci was specific to the SATIII probe.

In the case of co-localisation of lncSATIII and SAFB1, our results suggested that there was no co-localisation. However, this result was likely to be inaccurate as the SAFB1 staining pattern following RNA-FISH was different from SAFB1 staining alone (Figure 3.10 and Figure 3.22); therefore, no conclusions could be drawn this experiment. RNA staining was performed before ICC to avoid RNA degradation, suggesting that the change in SAFB staining pattern following RNA-FISH could be due to formamide in the hybridisation buffer for the RNA-FISH technique (section 2.10). Formamide destabilises existing nucleic acid complexes and allows the RNA probe to hybridise at lower temperatures, making it an essential part of the protocol. One explanation could be that as SAFB1 binds directly to RNA and to proteins that bind to RNA, the addition of formamide might have destabilised the interactions between SAFB1 and its binding partners, resulting in altered SAFB1 staining.

In contrast, SAFB2 was found to be colocalised with large lncSATIII transcripts on the microscopic images on both time points following stress. This corroborates with a previous study in which SAFB2 was found as an interacting partner of lncSATIII in a HSATIII ChIRP-MS [20]. However, statistical relevance of SAFB2/SATIII transcripts co-localisation between different timepoints was not picked up by our modular image analysis protocol due to the significant variance of the counts, despite the apparent changes seen by the naked eye the images. This could be explained by the need for more fine adjustments of the lncSATIII and SAFB nSBs recognition algorithm or the high lncSATIII background staining. Inaccurate identification of the nSBs (in any channel) could be troubleshot by re-training the Trainable Weka Segmentation Tool and/or excluding images with high debris content. Statistical significance for SATIII transcripts and SAFB2 co-localisation was only found between NHS and HSNR time points for small nSBs. This suggests a stronger association between SAFB2 and the SATIII transcripts or that SAFB2 is strongly associated with SATIII binding proteins (e.g., HSF1).

As SAFB proteins recruit RNA polymerase II to sites of transcription we examined whether SAFB proteins might act as regulators of lncSATIII transcription. To do this, we knocked down SAFB proteins using RNAi constructs, confirmed that SAFB1 and SAFB2 mRNA

levels were significantly decreased compared to the NTC controls at all time points, and measured lncSATIII using RT-qPCR. Our results show that no SATIII transcripts could be detected in the absence of stress regardless of protein knockdown condition. The levels lncSATIII transcripts started to increase immediately after heat induction; however, they were not statistically different from the NHS controls. We did observe a significant increase in lncSATIII transcription in the NTC treatment after the recovery period. Knockdown of SAFB1 did not significantly affect lncSATII transcription. Whereas SAFB2 knockdown significantly increased lncSATII transcription. Surprisingly, when both SAFB1 and SAFB2 were knocked down, lncSATIII fold change was similar to non-stressed levels, suggesting that the presence of SAFB1 mRNA might have a role in the regulation of lncSATIII transcription. Thus, our results suggest that both SAFB1 and SAFB2 mRNA play a role in regulating lncSATIII transcription.

Next, we investigated the distribution patterns of HSF1 and SAFB proteins following HS in human fibroblasts to test whether or not a primary non-transformed cell would show similar recruitment dynamic. To do this, fibroblast in cell culture were subjected to HS treatment followed by ICC to detect HSF1 and its co-localisation with SAFB family members. Our results show that fibroblasts are more competent in assembling nuclear stress bodies, as ~90% of the fibroblasts exhibited HSF1+ nSBs immediately after HS. In contrast, we observed HSF1+ nSBs only in ~16.5% of SHSY cells. Furthermore, after one hour of recovery, the percentage of HSF1+ nSBs in fibroblasts drops to ~68%, which is still considerably higher than the ~18.5% observed in our SHSY cells. These results are in accordance with Jolly et al. I [200], where the researchers observed that the older the cells become, the fewer HSF1+ nSBs are observed, possibly linking nSBs formation to senescence. In SAFB1, there was no co-localisation of endogenous SAFB1 with the HSF1+ nSBs at any time point. On the other hand, endogenous SAFB2 co-localized with the HSF1+ nSBs following HS and the amount of co-localisation increased even further during recovery.

Additionally, after recovery, the SAFB2+ nSBs were larger than in the previous time point and even larger than the HSF1+ nSBs. Thus, our results suggest that HSF1 is recruited first, followed by the recruitment of additional proteins, SAFB2 amongst them. Once the

stress is removed HSF1 returns to its diffuse distribution, whereas SAFB2 is still being recruited and remains at the nSBs for a longer time.

Moving on now to consider the transcription of lncSATIII RNA in human fibroblasts following stress. We examined the co-localisation of SAFB1 and SAFB2 with SATIII+ nSBs by FISH and immunofluorescence. Our results show that primary fibroblasts express lncSATIII transcripts immediately after HS in a small percentage of cells, increasing to almost all cells after recovery. Interestingly, we did not observe co-localisation of SAFB1 with the lncSATIII transcripts until after 2 hrs of recovery. On the other hand, SAFB2 co-localised with lncSATIII transcripts immediately after HS and SAFB2 aggregates persisted after recovery. This result may be explained by the fact that following HS, HFS1 binds to SATIII genomic regions, recruits RNA pol II to generate SATIII transcripts, which attract and bind a variety of RNA biding proteins, including SAFB2. As the thermal stress is withdrawn, a different set of proteins is recruited to these nSBs to aid in the stress recovery including SAFB1. This is supported by previous studies where SAFB2 was found to interact with lncSATIII transcripts with high score (2316668 score) amongst all proteins identified, whereas SAFB1, showed a dramatic lower enrichment score (197.9823 score)[20]. However, a note of caution is due here since these observations were based on microscopy images, and no measurements on size or number were performed. To develop a full picture of SAFB protein recruitment to SATIII transcription regions, these experiments should be repeated to quantify the co-localisation of SAFB1 and SAFB2 with SATIII and confirm whether the differences we observed are significant.

Next, we characterised the effects of the expression of SAFB1 mutants lacking specific functional domains and examined the effects on SAFB protein localisation/distribution following heat stress. Our results showed that deletion of either the SAP, HS, RRM or Gly-Rich domains in HeLa cells resulted in disruption of SAFB1 localisation and the formation of multiple aggregates throughout the nucleus compartment before and after heat stress induction. This suggests that these domains are not crucial for determining SAFB1's nuclear localisation but are needed to correct the protein recruitment. Interestingly, when both Gly-Rich and E/R-rich domains are deleted, SAFB1 is also expressed in the cytoplasm even though this mutant retained its NLS, an effect not described before in the literature, suggesting that interactions via Gly-Rich and E/R-rich domains are crucial for

its nuclear localisation in HeLa cells. On the other hand, when this mutant was expressed in a neuronally derived cell line (SH-SY5Y cells), the ΔE/R_Gly-rich mutant SAFB1 protein was excluded entirely from the nucleus before HS but transported back into it upon HS and HS recovery. More importantly, ΔE/R_Gly-rich mutant SAFB1 protein did not colocalise with HSF1 in the nucleus nor cytoplasm, in contrast to what can be observed in cells overexpressing WT SAFB1 protein. This means that without the E/R_Gly-rich domains, SAFB1 cannot remain in the nucleus unless stress is applied and that these domains are crucial for exogenous SAFB1 recruitment to HSF1 nSBs. Interestingly, the E/R and Gly-rich domains contain 6 out of 7 RG/RGG motifs present in the SAFB1 protein sequence. These RG/RGG motifs can be recognised by PRMTs and are susceptible to Arginine methylation.

Posttranslational modifications (PTMs) modulate the functions of proteins [389-391]. Various studies have reported several types of PTMs present in SAFB proteins, such as phosphorylation [243], acetylation [127], SUMOylation [44], and methylation [122]. It is hypothesised that these PTMs produce changes in SAFB protein-protein interactions, but the effects of recently reported modifications, such as methylation and sumoylation, have not been elucidated. Protein-protein interactions between SAFB family members may be important regulators for their recruitment to nSBs. Therefore, this study sought to determine whether methylation or sumoylation of SAFB1 and SAFB2 are involved in their recruitment to nSBs during the HS response.

Firstly, we examined the effects of pharmacological methylation inhibitors on the recruitment of HSF1, SAFB1, SAFB2 and SLTM. We treated HeLa cells with a pan-methylation inhibitor (AdOx) and PRMT5 inhibitor (GSK591) before HS, then used ICC to stain HSF1, SAFB1, SAFB2, and SLTM. Our results showed that inhibition of global methylation using AdOx significantly decreased the formation of large and small HSF1+ nSBs, even after recovery. However, this result can be due to histone methylation, as histone methylation status is a regulator of chromatin opening [392]. In addition, as AdOx inhibits methylation in a non-specific manner, its use may have inhibited histone methylation, making DNA less accessible for transcription factor binding. Therefore, the reduction in HSF1+ nSBs could be explained by HSF1 having less access to its DNA binding sites.

On the other hand, specific inhibition of PRMT5 (a major generator of symmetric di-methyl-arginine) via GSK591 treatment led to an increase in the number of cells expressing large and small HSF1+ nSBs immediately after HS. Additionally, following recovery, the number of small HSF1+ nSBs is comparable to the control group, whereas large HSF1+ nSBs are still significantly higher than in the control group. These results were then used to compare the effects of methylation inhibitors on nSB formation alone to the effects of methylation inhibitors on SAFB1 and SAFB2 recruitment.

In SAFB1, global methylation inhibition increased the number of cells expressing large SAFB1 clusters at all time points, whereas PRMT5 inhibition did not affect SAFB1 distribution compared to the control group. However, co-localisation between large HSF1 and SAFB1 was decreased immediately after HS. Small SAFB1 clusters increased under normal conditions (no HS) following AdOx treatment, but there was no effect at other timepoints. It has been reported in the literature that SAFB1 can be mono-methylated and di-methylated (asymmetric) in different RG/RGG motifs [22]. Arginine mono- and di-(asymmetric) methylation are mostly carried out by PRMT1, which could explain the effects of global methylation inhibition by AdOx on SAFB1dynamics during HS. The disruption of SAFB1 recruitment pattern upon AdOx and HS treatments suggests that mono- and symmetric di-methylation at arginine residues of SAFB1 are important for the correct function of SAFB1 following thermal stress. Furthermore, PRMT5 inhibition did not affect the formation of small SAFB1 clusters. Additionally, no co-localisation between small HSF1+ nSBs and SAFB1 was observed. These results show that by inhibiting global methylation, SAFB1 forms small clusters more readily in the absence of HS, and it forms large clusters at all time points, suggesting that methylation influences the distribution pattern of SAFB1. However, as the results were observed on cells where global methylation was inhibited, the effects might be due to the lack of access to DNA binding sites.

Turning now to the effects of global methylation inhibition on SAFB2, we found that the formation of SAFB2 large and small clusters is not affected by the inhibitor used at any time point. Studies have also reported that arginine residues of SAFB2 can be mono- and asymmetrically di- methylated [22]. This suggests that arginine methylation is not essential for SAFB2 recruitment to the nSBs upon HS. In contrast, specific PRMT5

inhibition decreases SAFB2 co-localisation with small HSF1+ nSBs immediately after HS. However, this effect is reversed after a recovery period. These results suggest that in SAFB2, the specific inhibition of symmetric di-methylation affects the recruitment of SAFB2 to small HSF1+ nSBs. It is important to note that the formation of HSF1+ nSBs was not affected by the inhibition of PRMT5, indicating that SAFB2 methylation by PRMT5 is important for the presence of SAFB2 at nSBs immediately after HS. However, whether PRMT5 effect on SAFB2 is a result of SAFB2 not being symmetrically di-methylated or an indirect effect of PRMT5 on other proteins remains unknown. In future investigations, it might be possible to perform SAFB2 immunoprecipitation followed by mass spectrometry for the identification of SDMA presence to confirm whether symmetric di-methylation is involved in SAFB2 recruitment to nSBs.

In the case of SLTM, neither large nor small SLTM puncta formation was affected by the methylation inhibitors at any time point. However, PRMT5 inhibition significantly reduced the co-localisation between large HSF1+ nSBs and SLTM after HS followed by recovery. These results suggest that similarly to SAFB2, the specific inhibition of methylation by PRMT5 alters the permanence of SLTM at nSBs, following recovery in this case.

The next step was to examine the effects of methylation by expressing mutant SAFB proteins in which RG/RGG motifs, sites of recognition for PRMTs [141, 393], were replaced by KG/KGG motifs. We generated and overexpressed the KG/KGG mutants of SAFB1 and SAFB2 in HeLa cells and observed that both mutants were still retained within the nuclear compartment, thus, not mimicking the effects seen with overexpression of the ΔE/R_Gly-rich mutant SAFB1. Moreover, fewer cells formed aggregates of overexpressed SAFB1 and SAFB2 immediately after HS than a wildtype control, suggesting that arginine methylation in SAFB proteins is involved in their recruitment to stress granules upon stress conditions. However, following recovery, there were no differences between either SAFB1 or SAFB2 KG/KGG mutant and control, suggesting that Arginine methylation on SAFB proteins is not essential SAFB1/2 during recovery periods.

SUMOylation is another PTM that may impact the recruitment of SAFB proteins to stress granules, and previous studies have shown that both SAFB1 and SAFB2 are SUMOylated on two lysine residues within the High Similarity Domain [44]. Therefore, we synthesised

SAFB1 and SAFB2 mutants in which the lysine on its SUMOylation sites was replaced by arginine, impairing SUMOylation. We observed that impairment SUMOylation is not crucial for SAFB1/2 translocation to the cytoplasm, as both mutated proteins were expressed in the nucleus. Moreover, overexpression of either SAFB1 or SAFB2 with non-functional SUMO sites resulted in fewer cells forming SAFB1/2 nuclear aggregates immediately after HS. However, this effect was not observed following 1 hr of recovery on overexpressed SAFB1 KG/KGG mutants, as the percentage of cells forming puncta were comparable to controls.

On the other hand, impairment of SAFB2 SUMOylation significantly decreased the ability of the mutated protein to form stress aggregates after recovery. Moreover, the impairment of SUMOylation sites in SAFB2 also significantly decreased the co-localisation of SAFB2 with HSF1+ nSBs following HS recovery. These results reinforce our hypothesis that SAFB2 has a different role on the HSR from SAFB1, such as binding miRNAs and/or regulate processing by microprocessor (discussed below), and it suggests that SAFB2 SUMOylation is important for the efficient recruitment of SAFB2 nSBs. Further research should be undertaken to investigate whether these effects are also present in endogenous SAFB proteins by using either chemical inhibition of SUMOylation/de-SUMOylation or knockdown of E1, E2 and E3 enzymes.

We conducted co-immunoprecipitation experiments followed by Tandem Mass Tagging spectrometry (TMT-MS) with SAFB specific antibodies to discover new interacting partners of SAFB proteins. Bioinformatics analysis of SAFB1/SAFB2-interacting complexes showed that 1072 proteins interact with SAFB1, 288 (24.7%) of them unique to SAFB1. Whereas 878 binding partners were found to bind to SAFB2, 94 (8.1%) were unique to SAFB2. The remaining 784 proteins were identified as SAFB1 and SAFB2 interacting partners, comprising 67.2% of the total number of proteins associated with SAFB1 and SAFB2. This finding suggests that both SAFB1 and SAFB2 interact with most of the proteins found, but they also have distinct interactions, which may explain the differences seen in their distribution patterns following HS. However, further experiments should confirm which of these different binding partners mediate those different functions.

Our findings are consistent with previously described functions of SAFB proteins as most SAFB1 interacting proteins are RNA binding proteins, with roles in the regulation of gene

expression (RBM 26 and 27, ELAVL1, LARP4, NSRP1, FLII), RNA processing (CPSF6, HNRNPA2B1, SYNCRIP, HNRNPR), splicing (EIF4A3, DHX15, SNRNP70), chromatin scaffolding (MATR3, AKAP13) and cell cycle (HP1BP3, CEP170, DHX9). In addition, ELAVL1 [319] and MATR3 [320] are known interactors of SAFB1; and, most SAFB2 interaction partners are involved in RNA metabolism, with roles in protein translation (PRKRA, Elongin-A), gene silencing (XPO1), splicing (SNRPD3, EIF2S1, SNRPD2, HNRNPF, RNPS1), RNA export (THOC1, CAPRIN1), RNA processing (HNRNPA0) and RNA binding protein (NUFIP1). Therefore, the analysis of SAFB protein interactions undertaken here has extended our knowledge of the specific SAFB1 and SAFB2 binding partners, as some of them had not been previously reported.

Furthermore, our interactome analysis identified Ago2 - Argonaute RISC catalytic component 2 (accession NM_012154) -, amongst the interacting partners of SAFB1 (B1/IgG fold change: 5.137, FDR: 0.02167) and SAFB2 (B2/IgG fold change: 3.598, FDR: 0.04052). Additionally, our co-immunoprecipitation experiments showed that SAFB1 and SAFB2 interact directly or indirectly with Ago2 proteins under non-stress conditions. Interestingly, when thermal stress is inducted, we observed decreased band density in the pulldown samples, indicating that SAFB1/2 interactions with Ago2 might change upon stress. Ago2 is a member of the RNA-induced silencing complex (RISC) with an essential role in small RNAs guided post-transcriptional gene silencing, including mRNA degradation and translational repression (reviewed in [343, 344]), which has been linked to the development of Huntington's disease [100]. These results implicate SAFB1 and SAFB2 in transcriptional regulation.

Moreover, previous studies reported that SAFB1 and SAFB2 interact with the Microprocessor Protein Complex (MPC) within the canonical pathway of miRNA biogenesis: Drosha and DGCR8. DGCR8 is the non-catalytic subunit of the MPC, required for the biding of doubled stranded RNA to the complex, whereas Drosha is a double-stranded RNA ribonuclease III that cleaves the pri-miRNA into a short hairpin structure termed pre-miRNA [333, 334]. Finally, the pre-miRNAs are exported to the cytoplasm, where Dicer and Argonaute (Ago1-4) will process it to its mature form [335]. Interestingly, SAFB1 and SAFB2 co-immunoprecipitated with Ago2, TNRC6A and Exportin-1, responsible for exporting Ago2 to the cytoplasm via TNRC6A [336].

Other proteins found to interact with SAFB1/SAFB2 that are also involved in the miRNA processing are MOV10 an RNA helicase required for gene silencing by RISC [337]; FMR1, which may have roles in mRNA trafficking between nucleus and cytoplasm, especially in the brain [338, 339]; ELAVL1, PRKRA, RBM4, SRRT, HNRNPA2B1, eIF6 and eIF4E. Additionally, SAFB1 and SAFB2 were found to interact with the pre-mRNA-processing factor 6 (PRP6), which is involved in the transactivation of the androgen receptor (AR) [340], and it is a component of the U4/U6-U5 tri-snRNP complex, as part of the spliceosome [341]; with the Serine/Arginine-related protein 53 (SRrp53 or RSRC1) which plays a role in alternative splicing and transcription regulation [342].

Having obtained the list of SAFB1 and SAFB2 interacting partners, we performed an Ingenuity Pathway Analysis (IPA) on our proteomic data to identify the pathways in which those interacting partners are involved. The analysis of canonical pathways suggested that both SAFB1 and SAFB2 are most enriched within the eIF2 signalling pathway (Eukaryotic Initiation Factor 2), with a positive Z-score indicating both proteins act towards the activation of the pathway. The eIF2 pathway restores homeostasis in response to stress as part of the Integrated Stress Response (ISR). Thus, the enrichment of SAFB proteins in the pathway is consistent with our results and previous studies placing SAFB 1/2 as part of the cellular response to stress and the ISR. Furthermore, we found several proteins within this canonical pathway interacted directly or indirectly with SAFB1 and SAFB2, including Ago2 and HNRNPA1, both involved in RNA processing, placing SAFB1/SAFB2 as a possible link between these biological processes.

Another pathway in which SAFB1 and SAFB2 were significantly enriched is the Unfolded Protein Response (UPR) pathway. The enrichment of SAFB1/2 proteins in this pathway is consistent with our results showing recruitment of SAFB2 to the HSF1+ nSBs/sites of SATIII transcript accumulation, especially because SAFB2 shows higher Z-scores and p-values than SAFB1 (which indicate the likelihood that the binding protein activates the pathway). Furthermore, SAFB1 and SAFB2 interact with Heat Shock Proteins in this pathway, including BIP, a member of the eIF2 Signalling pathway, constituting a link between the UPR and eIF2 signalling.

Interestingly, the Huntington's Disease Signalling pathway appeared as significantly enriched for SAFB proteins as well. SAFB proteins are highly expressed in the brain [356],

especially in the hippocampus [30], and they are abnormally expressed in Huntington's Disease [95]. More importantly, although Ago2 has not been included in the canonical pathway for Huntington's Disease, it has been recently linked to it as Ago2 accumulates in striatal neurons expressing mHTT protein [100]. Since we have shown that SAFB1 and SAFB2 interact with Ago2, this could explain the abnormal expression of SAFB1 in the nucleus and cytoplasm of striatal neurons in HD patients seen by our research group members [95].

Another significant aspect of the IPA network analyses of SAFB1 and SAFB2 is an extremely high enrichment of both proteins within the RNA metabolism cluster. Functional analysis by IPA generated networks showed that the top five networks in which SAFB1 was involved mediate RNA post-transcriptional modifications, reinforcing the hypothesis that SAFB1 is part of the RNA transcription machinery. SAFB2 is involved in a more diverse range of clusters, such as cancer/cell death and survival, cellular assembling and organisation, DNA replication/recombination and cell cycle, besides RNA post-transcriptional modifications. SAFB2 has recently been recognised as an accessory protein of the microprocessor, enabling the binding and processing of suboptimal Microprocessor substrates in clustered primary miRNA transcripts [51].

Considering the enrichment of SAFB1/SAFB2 in bio functions and disease analysis, the top ones were involved with were RNA/mRNA processing, splicing and alternative mRNA splicing, crucial for protein expression and protein variability; rRNA processing and modifications, involved in protein translation; protein metabolism (translation, synthesis, elongation, expression and termination); and molecular trafficking. Thus, this result reinforces once more the involvement of SAFB1 and SAFB2 within the RNA metabolism.

The purpose of the current study was to determine whether certain SAFB family members have specific roles within cell biology. We hypothesized that SAFB2 plays a more active role within the stress response than SAFB1 or SLTM, and that SAFB2 may also interact with a different group of binding partners from SAFB1 and therefore, show a different dynamic within the stress response.

Having discussed the main findings, the following conclusions can be drawn from the present study: antibodies raised against specific SAFB family members are essential to differentiate the unique functions of these proteins within cell metabolism and avoid misinterpreted conclusions; SAFB2 has a different role in the response to thermal stress from SAFB1, however, more research should be conducted to verify whether SAFB2 plays a more active role than SAFB1 within the stress response; post-translational modifications, specifically SUMOylation and methylation, may play a role in regulating the activity of SAFB1 and SAFB2 within the stress response; SAFB1 and SAFB2 plays a role within the Integrated Stress Response, the Unfolded protein Response and Huntington Disease pathway; they are both highly enriched within miRNA processing mechanisms and interact with Ago2 placing them as important regulators of the microprocessor; and that even though SAFB1 and SAFB2 share many binding partners, they have unique protein-protein interactions that may be responsible for the differences in their functions.

Chapter 07: References

1. Weighardt, F., et al., *A novel hnRNP protein (HAP/SAF-B) enters a subset of hnRNP complexes and relocates in nuclear granules in response to heat shock.* Journal of Cell Science, 1999. **112**(10): p. 1465-1476.
2. Garee, J.P. and S. Oesterreich, *SAFB1's multiple functions in biological control-lots still to be done!* J Cell Biochem, 2010. **109**(2): p. 312-9.
3. Norman, M., et al., *The increasing diversity of functions attributed to the SAFB family of RNA-/DNA-binding proteins.* Biochem J, 2016. **473**(23): p. 4271-4288.
4. Townson, S.M., et al., *SAFB2, a new scaffold attachment factor homolog and estrogen receptor corepressor.* J Biol Chem, 2003. **278**(22): p. 20059-68.
5. Chan, C.W., et al., *A novel member of the SAF (scaffold attachment factor)-box protein family inhibits gene expression and induces apoptosis.* Biochem J, 2007. **407**(3): p. 355-62.
6. Renz, A. and F.O. Fackelmayer, *Purification and molecular cloning of the scaffold attachment factor B (SAF-B), a novel human nuclear protein that specifically binds to S/MAR-DNA.* Nucleic Acids Res, 1996. **24**(5): p. 843-9.
7. Oesterreich, S., et al., *Novel nuclear matrix protein HET binds to and influences activity of the HSP27 promoter in human breast cancer cells.* J Cell Biochem, 1997. **67**(2): p. 275-86.
8. Colley, S.M., et al., *MET, a novel stimulator of estrogen-induced transcription isolated from mouse bone marrow.* Journal of Bone and Mineral Research, 2002. **17**: p. S153-S153.
9. Huo, X., et al., *The Nuclear Matrix Protein SAFB Cooperates with Major Satellite RNAs to Stabilize Heterochromatin Architecture Partially through Phase Separation.* Mol Cell, 2019.
10. Soldi, M. and T. Bonaldi, *The Proteomic Investigation of Chromatin Functional Domains Reveals Novel Synergisms among Distinct Heterochromatin Components.* Molecular & Cellular Proteomics, 2013. **12**(3): p. 764-780.
11. Altmeyer, M., et al., *The chromatin scaffold protein SAFB1 renders chromatin permissive for DNA damage signaling.* Mol Cell, 2013. **52**(2): p. 206-20.
12. Cockerill, P.N. and W.T. Garrard, *Chromosomal Loop Anchorage Sites Appear to Be Evolutionarily Conserved.* Febs Letters, 1986. **204**(1): p. 5-7.
13. Kipp, M., et al., *SAF-Box, a conserved protein domain that specifically recognizes scaffold attachment region DNA.* Molecular and Cellular Biology, 2000. **20**(20): p. 7480-7489.
14. Yamaguchi, A. and K. Takanashi, *FUS interacts with nuclear matrix-associated protein SAFB1 as well as Matrin3 to regulate splicing and ligand-mediated transcription.* Sci Rep, 2016. **6**: p. 35195.
15. Aly, M.K., et al., *Two distinct nuclear stress bodies containing different sets of RNA-binding proteins are formed with HSATIII architectural noncoding RNAs upon thermal stress exposure.* Biochemical and Biophysical Research Communications, 2019. **516**(2): p. 419-423.
16. Liu, H.W., et al., *The chromatin scaffold protein SAFB1 localizes SUMO-1 to the promoters of ribosomal protein genes to facilitate transcription initiation and splicing.* Nucleic Acids Res, 2015. **43**(7): p. 3605-13.
17. Maris, C., C. Dominguez, and F.H. Allain, *The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression.* FEBS J, 2005. **272**(9): p. 2118-31.
18. Oubridge, C., et al., *Crystal structure at 1.92 A resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin.* Nature, 1994. **372**(6505): p. 432-8.
19. Brenner, S. and J.H. Miller, *Encyclopedia of genetics*. 2002, San Diego, Calif. ; London: Academic.
20. Ninomiya, K., et al., *LncRNA-dependent nuclear stress bodies promote intron retention through SR protein phosphorylation.* EMBO J, 2020. **39**(3): p. e102729.
21. Nigg, E.A., *Nucleocytoplasmic transport: signals, mechanisms and regulation.* Nature, 1997. **386**(6627): p. 779-87.

22. Van Dusen, C.M., et al., *A Glycine-Rich Domain of hnRNP H/F Promotes Nucleocytoplasmic Shuttling and Nuclear Import through an Interaction with Transportin 1.* Molecular and Cellular Biology, 2010. **30**(10): p. 2552-2562.
23. Dingwall, C. and R.A. Laskey, *Nuclear Targeting Sequences - a Consensus.* Trends in Biochemical Sciences, 1991. **16**(12): p. 478-481.
24. Wang, Y., et al., *Coiled-coil networking shapes cell molecular machinery.* Mol Biol Cell, 2012. **23**(19): p. 3911-22.
25. Drakouli, S., et al., *Enhancer of rudimentary homologue interacts with scaffold attachment factor B at the nuclear matrix to regulate SR protein phosphorylation.* FEBS J, 2017. **284**(15): p. 2482-2500.
26. Weng, M.T., et al., *Evolutionarily conserved protein ERH controls CENP-E mRNA splicing and is required for the survival of KRAS mutant cancer cells.* Proc Natl Acad Sci U S A, 2012. **109**(52): p. E3659-67.
27. DuPont, B.R., et al., *Assignment of SAFB encoding Hsp27 ERE-TATA binding protein (HET)/scaffold attachment factor B (SAF-B) to human chromosome 19 band p13.* Cytogenet Cell Genet, 1997. **79**(3-4): p. 284-5.
28. Baltz, A.G., et al., *The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts.* Mol Cell, 2012. **46**(5): p. 674-90.
29. Coordinators, N.R., *Database Resources of the National Center for Biotechnology Information.* Nucleic Acids Res, 2017. **45**(D1): p. D12-D17.
30. Rivers, C., et al., *iCLIP identifies novel roles for SAFB1 in regulating RNA processing and neuronal function.* BMC Biol, 2015. **13**: p. 111.
31. Jiang, S., et al., *Scaffold attachment factor B2 (SAFB2)-null mice reveal non-redundant functions of SAFB2 compared with its paralog, SAFB1.* Dis Model Mech, 2015. **8**(9): p. 1121-7.
32. Mukhopadhyay, N.K., et al., *Scaffold attachment factor B1 regulates the androgen receptor in concert with the growth inhibitory kinase MST1 and the methyltransferase EZH2.* Oncogene, 2014. **33**(25): p. 3235-3245.
33. Ansari, K.I., et al., *HOXC6 Is Transcriptionally Regulated via Coordination of MLL Histone Methylase and Estrogen Receptor in an Estrogen Environment.* Journal of Molecular Biology, 2011. **411**(2): p. 334-349.
34. Debril, M.B., et al., *Scaffold attachment factor B1 directly interacts with nuclear receptors in living cells and represses transcriptional activity.* J Mol Endocrinol, 2005. **35**(3): p. 503-17.
35. Ivanova, M., et al., *Scaffold attachment factor B1 functions in development, growth, and reproduction.* Mol Cell Biol, 2005. **25**(8): p. 2995-3006.
36. Uhlen, M., et al., *Tissue-based map of the human proteome.* Science, 2015. **347**(6220).
37. Zhang, Z.L., et al., *A proteomic approach identifies SAFB-like transcription modulator (SLTM) as a bidirectional regulator of GLI family zinc finger transcription factors.* Journal of Biological Chemistry, 2019. **294**(14): p. 5549-5561.
38. Kalderon, D., et al., *A Short Amino-Acid Sequence Able to Specify Nuclear Location.* Cell, 1984. **39**(3): p. 499-509.
39. Higelin, J., et al., *FUS Mislocalization and Vulnerability to DNA Damage in ALS Patients Derived hiPSCs and Aging Motoneurons.* Front Cell Neurosci, 2016. **10**: p. 290.
40. Oesterreich, S., et al., *Tamoxifen-bound estrogen receptor (ER) strongly interacts with the nuclear matrix protein HET/SAF-B, a novel inhibitor of ER-mediated transactivation.* Molecular Endocrinology, 2000. **14**(3): p. 369-381.
41. Eymery, A., et al., *Heat shock factor 1 binds to and transcribes satellite II and III sequences at several pericentromeric regions in heat-shocked cells.* Exp Cell Res, 2010. **316**(11): p. 1845-55.
42. Dobrzycka, K.M., et al., *Disruption of scaffold attachment factor B1 leads to TBX2 up-regulation, lack of p19ARF induction, lack of senescence, and cell immortalization.* Cancer Res, 2006. **66**(16): p. 7859-63.

43. Hammerich-Hille, S., et al., *SAFB1 mediates repression of immune regulators and apoptotic genes in breast cancer cells.* J Biol Chem, 2010. **285**(6): p. 3608-16.
44. Garee, J.P., R. Meyer, and S. Oesterreich, *Co-repressor activity of scaffold attachment factor B1 requires sumoylation.* Biochem Biophys Res Commun, 2011. **408**(4): p. 516-22.
45. Hernandez-Hernandez, J.M., et al., *The Scaffold attachment factor b1 (Safb1) regulates myogenic differentiation by facilitating the transition of myogenic gene chromatin from a repressed to an activated state.* Nucleic Acids Res, 2013. **41**(11): p. 5704-16.
46. Chu, C., et al., *Systematic Discovery of Xist RNA Binding Proteins.* Cell, 2015. **161**(2): p. 404-416.
47. Townson, S.M., et al., *Structure-function analysis of the estrogen receptor alpha corepressor scaffold attachment factor-B1: identification of a potent transcriptional repression domain.* J Biol Chem, 2004. **279**(25): p. 26074-81.
48. Jiang, S., et al., *Scaffold attachment factor SAFB1 suppresses estrogen receptor alpha-mediated transcription in part via interaction with nuclear receptor corepressor.* Mol Endocrinol, 2006. **20**(2): p. 311-20.
49. Sergeant, K.A., et al., *Alternative RNA splicing complexes containing the scaffold attachment factor SAFB2.* J Cell Sci, 2007. **120**(Pt 2): p. 309-19.
50. Peidis, P., et al., *SAFB1 interacts with and suppresses the transcriptional activity of p53.* FEBS Lett, 2011. **585**(1): p. 78-84.
51. Hutter, K., et al., *SAFB2 Enables the Processing of Suboptimal Stem-Loop Structures in Clustered Primary miRNA Transcripts.* Mol Cell, 2020. **78**(5): p. 876-889 e6.
52. Lachapelle, S., S. Oesterreich, and M. Lebel, *The Werner syndrome helicase protein is required for cell proliferation, immortalization, and tumorigenesis in Scaffold attachment factor B1 deficient mice.* Aging (Albany NY), 2011. **3**(3): p. 277-90.
53. Hu, G., et al., *Functional Analysis of Human Hub Proteins and Their Interactors Involved in the Intrinsic Disorder-Enriched Interactions.* International Journal of Molecular Sciences, 2017. **18**(12).
54. Uversky, V.N., *Intrinsically Disordered Proteins and Their "Mysterious" (Meta)Physics.* Frontiers in Physics, 2019. **7**.
55. Dunker, A.K., et al., *Intrinsically disordered protein.* J Mol Graph Model, 2001. **19**(1): p. 26-59.
56. Uversky, V.N., *A protein-chameleon: conformational plasticity of alpha-synuclein, a disordered protein involved in neurodegenerative disorders.* J Biomol Struct Dyn, 2003. **21**(2): p. 211-34.
57. Uversky, V.N., *Intrinsic disorder in proteins associated with neurodegenerative diseases.* Front Biosci (Landmark Ed), 2009. **14**: p. 5188-238.
58. Santamaria, N., et al., *Intrinsic disorder in proteins involved in amyotrophic lateral sclerosis.* Cell Mol Life Sci, 2017. **74**(7): p. 1297-1318.
59. Wells, M., et al., *Structure of tumor suppressor p53 and its intrinsically disordered N-terminal transactivation domain.* Proc Natl Acad Sci U S A, 2008. **105**(15): p. 5762-7.
60. Mark, W.Y., et al., *Characterization of segments from the central region of BRCA1: an intrinsically disordered scaffold for multiple protein-protein and protein-DNA interactions?* J Mol Biol, 2005. **345**(2): p. 275-87.
61. Udan-Johns, M., et al., *Prion-like nuclear aggregation of TDP-43 during heat shock is regulated by HSP40/70 chaperones.* Hum Mol Genet, 2014. **23**(1): p. 157-70.
62. Song, M., et al., *Quantitative proteomic identification of the BRCA1 ubiquitination substrates.* J Proteome Res, 2011. **10**(11): p. 5191-8.
63. Lieberman-Aiden, E., et al., *Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome.* Science, 2009. **326**(5950): p. 289-293.
64. Nora, E.P., et al., *Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization.* Cell, 2017. **169**(5): p. 930-+.

65. Dixon, J.R., et al., *Topological domains in mammalian genomes identified by analysis of chromatin interactions.* Nature, 2012. **485**(7398): p. 376-380.
66. Shin, Y., et al., *Liquid Nuclear Condensates Mechanically Sense and Restructure the Genome (vol 175, pg 1481, 2018).* Cell, 2019. **176**(6).
67. Norwood, L.E., et al., *Conserved properties of HP1(Hs alpha).* Gene, 2004. **336**(1): p. 37-46.
68. Hashimoto, T., K. Matsuda, and M. Kawata, *Scaffold attachment factor B (SAFB)1 and SAFB2 cooperatively inhibit the intranuclear mobility and function of ERalpha.* J Cell Biochem, 2012. **113**(9): p. 3039-50.
69. Paggi, M.G. and A. Giordano, *Who is the boss in the Retinoblastoma family? The point of view of Rb2/p130, the little brother.* Cancer Research, 2001. **61**(12): p. 4651-4654.
70. Venkitaraman, A.R., *Cancer susceptibility and the functions of BRCA1 and BRCA2.* Cell, 2002. **108**(2): p. 171-182.
71. Lagier-Tourenne, C., M. Polymenidou, and D.W. Cleveland, *TDP-43 and FUS/TLS: emerging roles in RNA processing and neurodegeneration.* Hum Mol Genet, 2010. **19**(R1): p. R46-64.
72. Sergeant, K.A., et al., *Alternative RNA splicing complexes containing the scaffold attachment factor SAFB2.* Journal of Cell Science, 2007. **120**(2): p. 309-319.
73. Rappsilber, J., et al., *Large-scale proteomic analysis of the human spliceosome.* Genome Res, 2002. **12**(8): p. 1231-45.
74. Denegri, M., et al., *Human chromosomes 9, 12, and 15 contain the nucleation sites of stress-induced nuclear bodies.* Mol Biol Cell, 2002. **13**(6): p. 2069-79.
75. Finelli, P., et al., *Structural organization of multiple alphoid subsets coexisting on human chromosomes 1, 4, 5, 7, 9, 15, 18, and 19.* Genomics, 1996. **38**(3): p. 325-330.
76. Willard, H.F., *Chromosome-Specific Organization of Human Alpha Satellite DNA.* American Journal of Human Genetics, 1985. **37**(3): p. 524-532.
77. Chiodi, I., et al., *RNA recognition motif 2 directs the recruitment of SF2/ASF to nuclear stress bodies.* Nucleic Acids Res, 2004. **32**(14): p. 4127-36.
78. Valgardsdottir, R., et al., *Transcription of Satellite III non-coding RNAs is a general stress response in human cells.* Nucleic Acids Res, 2008. **36**(2): p. 423-34.
79. Lin, J.J., et al., *Identification of Proteins Binding to E-Box/Ku86 Sites and Function of the Tumor Suppressor SAFB1 in Transcriptional Regulation of the Human Xanthine Oxidoreductase Gene.* Journal of Biological Chemistry, 2008. **283**(44): p. 29681-29689.
80. Tuteja, R. and N. Tuteja, *Ku autoantigen: A multifunctional DNA-binding protein.* Critical Reviews in Biochemistry and Molecular Biology, 2000. **35**(1): p. 1-33.
81. Xu, P., et al., *Characterization of proteins binding to E-box/Ku86 sites and function of Ku86 in transcriptional regulation of the human xanthine oxidoreductase gene.* Journal of Biological Chemistry, 2004. **279**(16): p. 16057-16063.
82. Lee, Y.B., et al., *SAFB re-distribution marks steps of the apoptotic process.* Exp Cell Res, 2007. **313**(18): p. 3914-23.
83. Hikisz, P. and Z.M. Kilianska, *Puma, a critical mediator of cell death - one decade on from its discovery.* Cellular & Molecular Biology Letters, 2012. **17**(4): p. 646-669.
84. Wang, J., et al., *siRNA Suppression of NEDD9 Inhibits Proliferation and Enhances Apoptosis in Renal Cell Carcinoma.* Oncology Research, 2015. **22**(4): p. 219-224.
85. Liu, W., et al., *Osteoprotegerin Induces Apoptosis of Osteoclasts and Osteoclast Precursor Cells via the Fas/Fas Ligand Pathway.* Plos One, 2015. **10**(11).
86. Fan, S., et al., *BRCA1 inhibition of estrogen receptor signaling in transfected cells.* Science, 1999. **284**(5418): p. 1354-6.
87. Fan, S., et al., *BRCA1 inhibition of estrogen receptor signaling in transfected cells.* Science, 1999. **284**(5418): p. 1354-1356.
88. Oesterreich, S., *Scaffold attachment factors SAFB1 and SAFB2: Innocent bystanders or critical players in breast tumorigenesis?* J Cell Biochem, 2003. **90**(4): p. 653-61.

89. Oesterreich, S., et al., *Critical balance of scaffold attachment factor SAFB levels play important role in breast tumor suppression.* Breast Cancer Research and Treatment, 2002. **76**: p. S45-S45.
90. Oesterreich, S., et al., *High rates of loss of heterozygosity on chromosome 19p13 in human breast cancer.* British Journal of Cancer, 2001. **84**(4): p. 493-498.
91. Sutherland, L.C., et al., *RNA binding motif (RBM) proteins: a novel family of apoptosis modulators?* J Cell Biochem, 2005. **94**(1): p. 5-24.
92. Townson, S.M., et al., *HET/SAF-B overexpression causes growth arrest and multinuclearity and is associated with aneuploidy in human breast cancer.* Clinical Cancer Research, 2000. **6**(9): p. 3788-3796.
93. Takeuchi, A., et al., *Autoantibody to scaffold attachment factor B (SAFB): A novel connective tissue disease-related autoantibody associated with interstitial lung disease.* J Autoimmun, 2017. **76**: p. 101-107.
94. Ma, L., et al., *Scaffold attachment factor B suppresses HIV-1 infection of CD4(+) T cells by preventing binding of RNA polymerase II to HIV-1's long terminal repeat.* J Biol Chem, 2018. **293**(31): p. 12177-12185.
95. Buckner, N., et al., *Abnormal scaffold attachment factor 1 expression and localization in spinocerebellar ataxias and Huntington's chorea.* Brain Pathol, 2020.
96. Savas, J.N., et al., *Huntington's disease protein contributes to RNA-mediated gene silencing through association with Argonaute and P bodies.* Proc Natl Acad Sci U S A, 2008. **105**(31): p. 10820-5.
97. Savas, J.N., et al., *A role for huntington disease protein in dendritic RNA granules.* J Biol Chem, 2010. **285**(17): p. 13142-53.
98. Banez-Coronel, M., et al., *A pathogenic mechanism in Huntington's disease involves small CAG-repeated RNAs with neurotoxic activity.* PLoS Genet, 2012. **8**(2): p. e1002481.
99. Gennarino, V.A., et al., *Pumilio1 haploinsufficiency leads to SCA1-like neurodegeneration by increasing wild-type Ataxin1 levels.* Cell, 2015. **160**(6): p. 1087-98.
100. Pircs, K., et al., *Huntingtin Aggregation Impairs Autophagy, Leading to Argonaute-2 Accumulation and Global MicroRNA Dysregulation.* Cell Rep, 2018. **24**(6): p. 1397-1406.
101. Dong, X. and S. Cong, *The Emerging Role of microRNAs in Polyglutamine Diseases.* Front Mol Neurosci, 2019. **12**: p. 156.
102. Krauss, S. and B.O. Evert, *The Role of MicroRNAs in Spinocerebellar Ataxia Type 3.* J Mol Biol, 2019. **431**(9): p. 1729-1742.
103. Nitschke, L., et al., *miR760 regulates ATXN1 levels via interaction with its 5 ' untranslated region.* Genes & Development, 2020. **34**(17-18): p. 1147-1160.
104. Jolly, C., Y. Usson, and R.I. Morimoto, *Rapid and reversible relocalization of heat shock factor 1 within seconds to nuclear stress granules.* Proceedings of the National Academy of Sciences of the United States of America, 1999. **96**(12): p. 6769-6774.
105. Weighardt, F., et al., *A novel hnRNP protein (HAP/SAF-B) enters a subset of hnRNP complexes and relocates in nuclear granules in response to heat shock.* Journal of cell science, 1999. **112 (Pt 10)**: p. 1465-76.
106. Valgardsdottir, R., et al., *Transcription of Satellite III non-coding RNAs is a general stress response in human cells.* Nucleic Acids Research, 2008. **36**(2): p. 423-434.
107. Biamonti, G. and C. Vourc'h, *Nuclear stress bodies.* Cold Spring Harbor perspectives in biology, 2010. **2**(6): p. a000695.
108. Gomez-Pastor, R., E.T. Burchfiel, and D.J. Thiele, *Regulation of heat shock transcription factors and their roles in physiology and disease.* Nat Rev Mol Cell Biol, 2018. **19**(1): p. 4-19.
109. Landles, C. and G.P. Bates, *Huntingtin and the molecular pathogenesis of Huntington's disease - Fourth in molecular medicine review series.* Embo Reports, 2004. **5**(10): p. 958-963.
110. Choudhary, C., et al., *Lysine Acetylation Targets Protein Complexes and Co-Regulates Major Cellular Functions.* Science, 2009. **325**(5942): p. 834-840.

111. Gauci, S., et al., *Lys-N and Trypsin Cover Complementary Parts of the Phosphoproteome in a Refined SCX-Based Approach.* Analytical Chemistry, 2009. **81**(11): p. 4493-4501.
112. Beausoleil, S.A., et al., *A probability-based approach for high-throughput protein phosphorylation analysis and site localization.* Nature Biotechnology, 2006. **24**(10): p. 1285-1292.
113. Olsen, J.V., et al., *Global, in vivo, and site-specific phosphorylation dynamics in signaling networks.* Cell, 2006. **127**(3): p. 635-648.
114. Dephoure, N., et al., *A quantitative atlas of mitotic phosphorylation.* Proceedings of the National Academy of Sciences of the United States of America, 2008. **105**(31): p. 10762-10767.
115. Olsen, J.V., et al., *Quantitative Phosphoproteomics Reveals Widespread Full Phosphorylation Site Occupancy During Mitosis.* Science Signaling, 2010. **3**(104).
116. Hendriks, I.A., et al., *Uncovering global SUMOylation signaling networks in a site-specific manner.* Nature Structural & Molecular Biology, 2014. **21**(10): p. 927-936.
117. Impens, F., et al., *Mapping of SUMO sites and analysis of SUMOylation changes induced by external stimuli.* Proceedings of the National Academy of Sciences of the United States of America, 2014. **111**(34): p. 12432-12437.
118. Sutinen, P., et al., *Sumoylation Modulates the Transcriptional Activity of Androgen Receptor in a Target Gene and Pathway Selective Manner.* Endocrine Reviews, 2014. **35**(3).
119. Cubenas-Potts, C., et al., *Identification of SUMO-2/3-modified proteins associated with mitotic chromosomes.* Proteomics, 2015. **15**(4): p. 763-772.
120. Xiao, Z.Y., et al., *System-wide Analysis of SUMOylation Dynamics in Response to Replication Stress Reveals Novel Small Ubiquitin-like Modified Target Proteins and Acceptor Lysines Relevant for Genome Stability.* Molecular & Cellular Proteomics, 2015. **14**(5): p. 1419-1434.
121. Hendriks, I.A., et al., *Site-specific mapping of the human SUMO proteome reveals co-modification with phosphorylation.* Nature Structural & Molecular Biology, 2017. **24**(3): p. 325-+.
122. Guo, A., et al., *Immunoaffinity enrichment and mass spectrometry analysis of protein methylation.* Mol Cell Proteomics, 2014. **13**(1): p. 372-87.
123. Drazic, A., et al., *The world of protein acetylation.* Biochimica Et Biophysica Acta-Proteins and Proteomics, 2016. **1864**(10): p. 1372-1401.
124. Starheim, K.K., K. Gevaert, and T. Arnesen, *Protein N-terminal acetyltransferases: when the start matters.* Trends in Biochemical Sciences, 2012. **37**(4): p. 152-161.
125. Baeza, J., M.J. Smallegan, and J.M. Denu, *Mechanisms and Dynamics of Protein Acetylation in Mitochondria.* Trends in Biochemical Sciences, 2016. **41**(3): p. 231-244.
126. Ardito, F., et al., *The crucial role of protein phosphorylation in cell signaling and its use as targeted therapy.* International Journal of Molecular Medicine, 2017. **40**(2): p. 271-280.
127. Nayler, O., et al., *SAF-B protein couples transcription and pre-mRNA splicing to SAR/MAR elements.* Nucleic Acids Res, 1998. **26**(15): p. 3542-9.
128. Zhou, H.J., et al., *Toward a Comprehensive Characterization of a Human Cancer Cell Phosphoproteome.* Journal of Proteome Research, 2013. **12**(1): p. 260-271.
129. Bian, Y.Y., et al., *An enzyme assisted RP-RPLC approach for in-depth analysis of human liver phosphoproteome.* Journal of Proteomics, 2014. **96**: p. 253-262.
130. Carrascal, M., et al., *Phosphorylation Analysis of Primary Human T Lymphocytes Using Sequential IMAC and Titanium Oxide Enrichment.* Journal of Proteome Research, 2008. **7**(12): p. 5167-5176.
131. Oppermann, F.S., et al., *Large-scale Proteomics Analysis of the Human Kinome.* Molecular & Cellular Proteomics, 2009. **8**(7): p. 1751-1764.
132. Rigbolt, K.T.G., et al., *System-Wide Temporal Characterization of the Proteome and Phosphoproteome of Human Embryonic Stem Cell Differentiation.* Science Signaling, 2011. **4**(164).

133. Mahajan, R., et al., *A small ubiquitin-related polypeptide involved in targeting RanGAP1 to nuclear pore complex protein RanBP2.* Cell, 1997. **88**(1): p. 97-107.
134. Johnson, E.S., *Protein modification by SUMO.* Annual Review of Biochemistry, 2004. **73**: p. 355-382.
135. Yeh, E.T.H., *SUMOylation and De-SUMOylation: Wrestling with Life's Processes.* Journal of Biological Chemistry, 2009. **284**(13): p. 8223-8227.
136. Alabert, C., J.N. Bianco, and P. Pasero, *Differential regulation of homologous recombination at DNA breaks and replication forks by the Mrc1 branch of the S-phase checkpoint.* Embo Journal, 2009. **28**(8): p. 1131-1141.
137. Cai, L., et al., *Proteome-wide Mapping of Endogenous SUMOylation Sites in Mouse Testis.* Mol Cell Proteomics, 2017. **16**(5): p. 717-727.
138. Hendriks, I.A., et al., *SUMO-2 Orchestrates Chromatin Modifiers in Response to DNA Damage.* Cell Reports, 2015. **10**(10): p. 1778-1791.
139. Biggar, K.K. and S.S.C. Li, *Non-histone protein methylation as a regulator of cellular signalling and function.* Nature Reviews Molecular Cell Biology, 2015. **16**(1): p. 5-17.
140. Stewart, M.D., J.W. Li, and J.M. Wong, *Relationship between histone H3 lysine 9 methylation, transcription repression, and heterochromatin protein 1 recruitment.* Molecular and Cellular Biology, 2005. **25**(7): p. 2525-2538.
141. Bedford, M.T. and S.G. Clarke, *Protein arginine methylation in mammals: who, what, and why.* Mol Cell, 2009. **33**(1): p. 1-13.
142. Blanc, R.S. and S. Richard, *Arginine Methylation: The Coming of Age.* Mol Cell, 2017. **65**(1): p. 8-24.
143. Bedford, M.T. and S. Richard, *Arginine methylation an emerging regulator of protein function.* Mol Cell, 2005. **18**(3): p. 263-72.
144. Boisvert, F.M., C.A. Chenard, and S. Richard, *Protein interfaces in signaling regulated by arginine methylation.* Sci STKE, 2005. **2005**(271): p. re2.
145. Wei, H., et al., *Protein arginine methylation of non-histone proteins and its role in diseases.* Cell Cycle, 2014. **13**(1): p. 32-41.
146. Zhang, X., Y. Huang, and X. Shi, *Emerging roles of lysine methylation on non-histone proteins.* Cell Mol Life Sci, 2015. **72**(22): p. 4257-72.
147. Abu-Farha, M., et al., *Proteomic analyses of the SMYD family interactomes identify HSP90 as a novel target for SMYD2.* Journal of Molecular Cell Biology, 2011. **3**(5): p. 301-308.
148. Cho, H.S., et al., *Enhanced HSP70 lysine methylation promotes proliferation of cancer cells through activation of Aurora kinase B.* Nature Communications, 2012. **3**.
149. Li, L.B., I. Vorobyov, and T.W. Allen, *The Different Interactions of Lysine and Arginine Side Chains with Lipid Membranes.* Journal of Physical Chemistry B, 2013. **117**(40): p. 11906-11920.
150. Di Lorenzo, A. and M.T. Bedford, *Histone arginine methylation.* Febs Letters, 2011. **585**(13): p. 2024-2031.
151. Buuh, Z.Y., Z. Lyu, and R.E. Wang, *Interrogating the Roles of Post-Translational Modifications of Non-Histone Proteins.* Journal of Medicinal Chemistry, 2018. **61**(8): p. 3239-3252.
152. Yang, Y. and M.T. Bedford, *Protein arginine methyltransferases and cancer.* Nat Rev Cancer, 2013. **13**(1): p. 37-50.
153. Shi, Y., et al., *Histone demethylation mediated by the nuclear amine oxidase homolog LSD1.* Cell, 2004. **119**(7): p. 941-53.
154. Wang, Y., et al., *Human PAD4 regulates histone arginine methylation levels via demethylimination.* Science, 2004. **306**(5694): p. 279-283.
155. Chang, B.S., et al., *JMJD6 is a histone arginine demethylase.* Science, 2007. **318**(5849): p. 444-447.
156. Najbauer, J., et al., *Peptides with Sequences Similar to Glycine, Arginine-Rich Motifs in Proteins Interacting with Rna Are Efficiently Recognized by Methyltransferase(S) Modifying*

Arginine in Numerous Proteins. Journal of Biological Chemistry, 1993. **268**(14): p. 10501-10509.
157. Zakrzewicz, D., et al., *Protein Arginine Methyltransferases (PRMTs): Promising Targets for the Treatment of Pulmonary Disorders.* International Journal of Molecular Sciences, 2012. **13**(10): p. 12383-12400.
158. Corley, S.M. and J.E. Gready, *Identification of the RGG box motif in Shadoo: RNA-binding and signaling roles?* Bioinform Biol Insights, 2008. **2**: p. 383-400.
159. Kiledjian, M. and G. Dreyfuss, *Primary Structure and Binding-Activity of the Hnrnp U-Protein - Binding Rna through Rgg Box.* Embo Journal, 1992. **11**(7): p. 2655-2664.
160. Thandapani, P., et al., *Defining the RGG/RG motif.* Mol Cell, 2013. **50**(5): p. 613-23.
161. Wan, C., et al., *Panorama of ancient metazoan macromolecular complexes.* Nature, 2015. **525**(7569): p. 339-44.
162. Rajyaguru, P., M.P. She, and R. Parker, *Scd6 Targets eIF4G to Repress Translation: RGG Motif Proteins as a Class of eIF4G-Binding Proteins.* Molecular Cell, 2012. **45**(2): p. 244-254.
163. Hadian, K., et al., *Identification of a Heterogeneous Nuclear Ribonucleoprotein-recognition Region in the HIV Rev Protein.* Journal of Biological Chemistry, 2009. **284**(48): p. 33384-33391.
164. Kusakawa, T., et al., *Functional interaction of hepatitis C virus NS5B with nucleolin GAR domain.* Journal of Biochemistry, 2007. **141**(6): p. 917-927.
165. Boisvert, F.M., et al., *Symmetrical dimethylarginine methylation is required for the localization of SMN in Cajal bodies and pre-mRNA splicing.* Journal of Cell Biology, 2002. **159**(6): p. 957-969.
166. Smith, W.A., et al., *Arginine methylation of RNA helicase a determines its subcellular localization.* J Biol Chem, 2004. **279**(22): p. 22795-8.
167. Boisvert, F.M., et al., *A proteomic analysis of arginine-methylated protein complexes.* Mol Cell Proteomics, 2003. **2**(12): p. 1319-30.
168. Boisvert, F.M., et al., *Arginine methylation of MRE11 by PRMT1 is required for DNA damage checkpoint control.* Genes Dev, 2005. **19**(6): p. 671-6.
169. Boisvert, F.M., et al., *The GAR motif of 53BP1 is arginine methylated by PRMT1 and is necessary for 53BP1 DNA binding activity.* Cell Cycle, 2005. **4**(12): p. 1834-41.
170. Boisvert, F.M., et al., *Methylation of MRE11 regulates its nuclear compartmentalization.* Cell Cycle, 2005. **4**(7): p. 981-9.
171. Oh, J. and L.S. Symington, *Role of the Mre11 Complex in Preserving Genome Integrity.* Genes, 2018. **9**(12).
172. Rocco, A.G., et al., *Characterization of the protein unfolding processes induced by urea and temperature.* Biophysical Journal, 2008. **94**(6): p. 2241-2251.
173. Richter, K., M. Haslbeck, and J. Buchner, *The heat shock response: life on the verge of death.* Mol Cell, 2010. **40**(2): p. 253-66.
174. Pirkkala, L., P. Nykanen, and L. Sistonen, *Roles of the heat shock transcription factors in regulation of the heat shock response and beyond.* FASEB J, 2001. **15**(7): p. 1118-31.
175. Bascos, N.A.D. and S.J. Landry, *A History of Molecular Chaperone Structures in the Protein Data Bank.* Int J Mol Sci, 2019. **20**(24).
176. Mahat, D.B., et al., *Mammalian Heat Shock Response and Mechanisms Underlying Its Genome-wide Transcriptional Regulation.* Molecular Cell, 2016. **62**(1): p. 63-78.
177. Neueder, A., et al., *Novel isoforms of heat shock transcription factor 1, HSF1gammaalpha and HSF1gammabeta, regulate chaperone protein gene transcription.* J Biol Chem, 2014. **289**(29): p. 19894-906.
178. Jolly, C., et al., *In vivo binding of active heat shock transcription factor 1 to human chromosome 9 heterochromatin during stress.* Journal of Cell Biology, 2002. **156**(5): p. 775-781.

179. Jolly, C., et al., *Stress-induced transcription of satellite III repeats.* J Cell Biol, 2004. **164**(1): p. 25-33.
180. Goenka, A., et al., *Human satellite-III non-coding RNAs modulate heat-shock-induced transcriptional repression.* J Cell Sci, 2016. **129**(19): p. 3541-3552.
181. Lindquist, S. and E.A. Craig, *The Heat-Shock Proteins.* Annual Review of Genetics, 1988. **22**: p. 631-677.
182. Akerfelt, M., R.I. Morimoto, and L. Sistonen, *Heat shock factors: integrators of cell stress, development and lifespan.* Nat Rev Mol Cell Biol, 2010. **11**(8): p. 545-55.
183. Ali, A., et al., *HSP90 interacts with and regulates the activity of heat shock factor 1 in Xenopus oocytes.* Molecular and Cellular Biology, 1998. **18**(9): p. 4949-4960.
184. Zou, J.Y., et al., *Repression of heat shock transcription factor HSF1 activation by HSP90 (HSP90 complex) that forms a stress-sensitive complex with HSF1.* Cell, 1998. **94**(4): p. 471-480.
185. Jolly, C., et al., *HSF1 transcription factor concentrates in nuclear foci during heat shock: relationship with transcription sites.* J Cell Sci, 1997. **110 (Pt 23)**: p. 2935-41.
186. Mercier, P.A., N.A. Winegarden, and J.T. Westwood, *Human heat shock factor 1 is predominantly a nuclear protein before and after heat stress.* Journal of Cell Science, 1999. **112**(16): p. 2765-2774.
187. Sandqvist, A., et al., *Heterotrimerization of heat-shock factors 1 and 2 provides a transcriptional switch in response to distinct stimuli.* Mol Biol Cell, 2009. **20**(5): p. 1340-7.
188. Guettouche, T., et al., *Analysis of phosphorylation of human heat shock factor 1 in cells experiencing a stress.* BMC Biochem, 2005. **6**: p. 4.
189. Cotto, J., S. Fox, and R. Morimoto, *HSF1 granules: a novel stress-induced nuclear compartment of human cells.* J Cell Sci, 1997. **110 (Pt 23)**: p. 2925-34.
190. Sengupta, S., R. Parihar, and S. Ganesh, *Satellite III non-coding RNAs show distinct and stress-specific patterns of induction.* Biochem Biophys Res Commun, 2009. **382**(1): p. 102-7.
191. Denegri, M., et al., *Stress-induced nuclear bodies are sites of accumulation of pre-mRNA processing factors.* Molecular Biology of the Cell, 2001. **12**(11): p. 3502-3514.
192. Metz, A., et al., *A key role for stress-induced satellite III transcripts in the relocalization of splicing factors into nuclear stress granules.* J Cell Sci, 2004. **117**(Pt 19): p. 4551-8.
193. Bukau, B., J. Weissman, and A. Horwich, *Molecular chaperones and protein quality control.* Cell, 2006. **125**(3): p. 443-451.
194. Lackie, R.E., et al., *The Hsp70/Hsp90 Chaperone Machinery in Neurodegenerative Diseases.* Front Neurosci, 2017. **11**: p. 254.
195. Mayer, M.P. and B. Bukau, *Hsp70 chaperones: cellular functions and molecular mechanism.* Cell Mol Life Sci, 2005. **62**(6): p. 670-84.
196. Ross, R.A., B.A. Spengler, and J.L. Biedler, *Coordinate morphological and biochemical interconversion of human neuroblastoma cells.* J Natl Cancer Inst, 1983. **71**(4): p. 741-7.
197. Biedler, J.L., et al., *Multiple neurotransmitter synthesis by human neuroblastoma cell lines and clones.* Cancer Res, 1978. **38**(11 Pt 1): p. 3751-7.
198. Dewhirst, M.W., et al., *Basic principles of thermal dosimetry and thermal thresholds for tissue damage from hyperthermia.* Int J Hyperthermia, 2003. **19**(3): p. 267-94.
199. Yarmolenko, P.S., et al., *Thresholds for thermal damage to normal tissues: an update.* Int J Hyperthermia, 2011. **27**(4): p. 320-43.
200. Jolly, C., et al., *HSF1 transcription factor concentrates in nuclear foci during heat shock: relationship with transcription sites.* Journal of Cell Science, 1997. **110**: p. 2935-2941.
201. Chiodi, I., et al., *Structure and dynamics of hnRNP-labelled nuclear bodies induced by stress treatments.* J Cell Sci, 2000. **113 (Pt 22)**: p. 4043-53.
202. Rizzi, N., et al., *Transcriptional activation of a constitutive heterochromatic domain of the human genome in response to heat shock.* Mol Biol Cell, 2004. **15**(2): p. 543-51.

203. Schindelin, J., et al., *Fiji: an open-source platform for biological-image analysis.* Nature Methods, 2012. **9**(7): p. 676-682.
204. Rueden, C.T., et al., *ImageJ2: ImageJ for the next generation of scientific image data.* Bmc Bioinformatics, 2017. **18**.
205. Neerad Phansalkar, S.M., Ashish Sabale, Madhuri Joshi, *Adaptive local thresholding for detection of nuclei in diversity stained cytology images.* 2011 International Conference on Communications and Signal Processing, 2011.
206. Arganda-Carreras, I., et al., *Trainable Weka Segmentation: a machine learning tool for microscopy pixel classification.* Bioinformatics, 2017. **33**(15): p. 2424-2426.
207. Legland, D., I. Arganda-Carreras, and P. Andrey, *MorphoLibJ: integrated library and plugins for mathematical morphology with ImageJ.* Bioinformatics, 2016. **32**(22): p. 3532-3534.
208. Welch, W.J., et al., *Response of Mammalian-Cells to Metabolic Stress - Changes in Cell Physiology and Structure-Function of Stress Proteins.* Current Topics in Microbiology and Immunology, 1991. **167**: p. 31-55.
209. Jolly, C. and R.I. Morimoto, *Stress and the cell nucleus: Dynamics of gene expression and structural reorganization.* Gene Expression, 1999. **7**(4-6): p. 261-270.
210. Pirkkala, L., P. Nykanen, and L. Sistonen, *Roles of the heat shock transcription factors in regulation of the heat shock response and beyond.* Faseb Journal, 2001. **15**(7): p. 1118-1131.
211. Weids, A.J., et al., *Distinct stress conditions result in aggregation of proteins with similar properties.* Scientific Reports, 2016. **6**.
212. Hong, E., et al., *Unravelling the RNA-Binding Properties of SAFB Proteins in Breast Cancer Cells.* Biomed Res Int, 2015. **2015**: p. 395816.
213. Uhlen, M., et al., *Proteomics. Tissue-based map of the human proteome.* Science, 2015. **347**(6220): p. 1260419.
214. Biamonti, G. and C. Vourc'h, *Nuclear stress bodies.* Cold Spring Harb Perspect Biol, 2010. **2**(6): p. a000695.
215. Lee, M., et al., *RRP1B is a metastasis modifier that regulates the expression of alternative mRNA isoforms through interactions with SRSF1.* Oncogene, 2014. **33**(14): p. 1818-1827.
216. Ding, F.J., et al., *Oxymatrine inhibits microglia activation via HSP60-TLR4 signaling.* Biomedical Reports, 2016. **5**(5): p. 623-628.
217. Djamali, A., et al., *Nox2 and Cyclosporine-Induced Renal Hypoxia.* Transplantation, 2016. **100**(6): p. 1198-1210.
218. Yang, X., et al., *5-methylcytosine promotes mRNA export - NSUN2 as the methyltransferase and ALYREF as an m(5)C reader.* Cell Res, 2017. **27**(5): p. 606-625.
219. Zhang, R., et al., *Neuroprotective effect of heat shock protein 60 on matrine-suppressed microglial activation.* Experimental and Therapeutic Medicine, 2017. **14**(2): p. 1832-1836.
220. Fontenete, S., et al., *Prediction of melting temperatures in fluorescence in situ hybridization (FISH) procedures using thermodynamic models.* Crit Rev Biotechnol, 2016. **36**(3): p. 566-77.
221. Harvey, R., et al., *Post-transcriptional control of gene expression following stress: the role of RNA-binding proteins.* Biochem Soc Trans, 2017. **45**(4): p. 1007-14.
222. Hawes, J.W., et al., *Evaluation of methods for sequence analysis of highly repetitive DNA templates.* J Biomol Tech, 2006. **17**(2): p. 138-44.
223. Denegri, M., et al., *Human chromosomes 9, 12, and 15 contain the nucleation sites of stress-induced nuclear bodies.* Molecular biology of the cell, 2002. **13**(6): p. 2069-79.
224. Jolly, C., et al., *In vivo binding of active heat shock transcription factor 1 to human chromosome 9 heterochromatin during stress.* The Journal of cell biology, 2002. **156**(5): p. 775-81.
225. Goenka, A., et al., *Human satellite-III non-coding RNAs modulate heat shock-induced transcriptional repression.* J Cell Sci, 2016.
226. Mann, M. and O.N. Jensen, *Proteomic analysis of post-translational modifications.* Nat Biotechnol, 2003. **21**(3): p. 255-61.

227. Deribe, Y.L., T. Pawson, and I. Dikic, *Post-translational modifications in signal integration.* Nat Struct Mol Biol, 2010. **17**(6): p. 666-72.
228. Khoury, G.A., R.C. Baliban, and C.A. Floudas, *Proteome-wide post-translational modification statistics: frequency analysis and curation of the swiss-prot database.* Sci Rep, 2011. **1**.
229. Beausoleil, S.A., et al., *A probability-based approach for high-throughput protein phosphorylation analysis and site localization.* Nat Biotechnol, 2006. **24**(10): p. 1285-92.
230. Olsen, J.V., et al., *Global, in vivo, and site-specific phosphorylation dynamics in signaling networks.* Cell, 2006. **127**(3): p. 635-48.
231. Cantin, G.T., et al., *Combining protein-based IMAC, peptide-based IMAC, and MudPIT for efficient phosphoproteomic analysis.* J Proteome Res, 2008. **7**(3): p. 1346-51.
232. Carrascal, M., et al., *Phosphorylation analysis of primary human T lymphocytes using sequential IMAC and titanium oxide enrichment.* J Proteome Res, 2008. **7**(12): p. 5167-76.
233. Daub, H., et al., *Kinase-selective enrichment enables quantitative phosphoproteomics of the kinome across the cell cycle.* Mol Cell, 2008. **31**(3): p. 438-48.
234. Dephoure, N., et al., *A quantitative atlas of mitotic phosphorylation.* Proc Natl Acad Sci U S A, 2008. **105**(31): p. 10762-7.
235. Choudhary, C., et al., *Lysine acetylation targets protein complexes and co-regulates major cellular functions.* Science, 2009. **325**(5942): p. 834-40.
236. Gauci, S., et al., *Lys-N and trypsin cover complementary parts of the phosphoproteome in a refined SCX-based approach.* Anal Chem, 2009. **81**(11): p. 4493-501.
237. Oppermann, F.S., et al., *Large-scale proteomics analysis of the human kinome.* Mol Cell Proteomics, 2009. **8**(7): p. 1751-64.
238. Olsen, J.V., et al., *Quantitative phosphoproteomics reveals widespread full phosphorylation site occupancy during mitosis.* Sci Signal, 2010. **3**(104): p. ra3.
239. Rigbolt, K.T., et al., *System-wide temporal characterization of the proteome and phosphoproteome of human embryonic stem cell differentiation.* Sci Signal, 2011. **4**(164): p. rs3.
240. Bienvenut, W.V., et al., *Comparative large scale characterization of plant versus mammal proteins reveals similar and idiosyncratic N-alpha-acetylation features.* Mol Cell Proteomics, 2012. **11**(6): p. M111.015131.
241. Van Damme, P., et al., *N-terminal acetylome analyses and functional insights of the N-terminal acetyltransferase NatB.* Proc Natl Acad Sci U S A, 2012. **109**(31): p. 12449-54.
242. Zhou, H., et al., *Toward a comprehensive characterization of a human cancer cell phosphoproteome.* J Proteome Res, 2013. **12**(1): p. 260-71.
243. Bian, Y., et al., *An enzyme assisted RP-RPLC approach for in-depth analysis of human liver phosphoproteome.* J Proteomics, 2014. **96**: p. 253-62.
244. Hendriks, I.A., et al., *Uncovering global SUMOylation signaling networks in a site-specific manner.* Nat Struct Mol Biol, 2014. **21**(10): p. 927-36.
245. Hendriks, I.A., et al., *SUMO-2 Orchestrates Chromatin Modifiers in Response to DNA Damage.* Cell Rep, 2015. **10**(10): p. 1778-1791.
246. Xiao, Z., et al., *System-wide Analysis of SUMOylation Dynamics in Response to Replication Stress Reveals Novel Small Ubiquitin-like Modified Target Proteins and Acceptor Lysines Relevant for Genome Stability.* Mol Cell Proteomics, 2015. **14**(5): p. 1419-34.
247. Hendriks, I.A., et al., *Site-specific mapping of the human SUMO proteome reveals co-modification with phosphorylation.* Nat Struct Mol Biol, 2017. **24**(3): p. 325-336.
248. Gong, F.D. and K.M. Miller, *Histone methylation and the DNA damage response.* Mutation Research-Reviews in Mutation Research, 2019. **780**: p. 37-47.
249. Urulangodi, M. and A. Mohanty, *DNA damage response and repair pathway modulation by non-histone protein methylation: implications in neurodegeneration.* J Cell Commun Signal, 2019.

250. Rossetto, D., N. Avvakumov, and J. Cote, *Histone phosphorylation: a chromatin modification involved in diverse nuclear events.* Epigenetics, 2012. **7**(10): p. 1098-108.
251. Biggar, K.K. and S.S. Li, *Non-histone protein methylation as a regulator of cellular signalling and function.* Nat Rev Mol Cell Biol, 2015. **16**(1): p. 5-17.
252. Gary, J.D. and S. Clarke, *RNA and protein interactions modulated by protein arginine methylation.* Progress in Nucleic Acid Research and Molecular Biology, Vol 61, 1998. **61**: p. 65-131.
253. Hu, H., et al., *Small Molecule Inhibitors of Protein Arginine Methyltransferases.* Expert Opin Investig Drugs, 2016. **25**(3): p. 335-58.
254. Poulard, C., L. Corbo, and M. Le Romancer, *Protein arginine methylation/demethylation and cancer.* Oncotarget, 2016. **7**(41): p. 67532-67550.
255. Wesche, J., et al., *Protein arginine methylation: a prominent modification and its demethylation.* Cell Mol Life Sci, 2017. **74**(18): p. 3305-3315.
256. Hyun, K., et al., *Writing, erasing and reading histone lysine methylations.* Exp Mol Med, 2017. **49**(4): p. e324.
257. Yang, Y., et al., *PRMT9 is a type II methyltransferase that methylates the splicing factor SAP145.* Nat Commun, 2015. **6**: p. 6428.
258. Feng, Y., et al., *Mammalian protein arginine methyltransferase 7 (PRMT7) specifically targets RXR sites in lysine- and arginine-rich regions.* J Biol Chem, 2013. **288**(52): p. 37010-25.
259. Tang, J., et al., *PRMT1 is the predominant type I protein arginine methyltransferase in mammalian cells.* J Biol Chem, 2000. **275**(11): p. 7723-30.
260. Herrmann, F., et al., *Dynamics of human protein arginine methyltransferase 1(PRMT1) in vivo.* J Biol Chem, 2005. **280**(45): p. 38005-10.
261. Liu, Q. and G. Dreyfuss, *In vivo and in vitro arginine methylation of RNA-binding proteins.* Mol Cell Biol, 1995. **15**(5): p. 2800-8.
262. Herrmann, F., et al., *Arginine methylation of scaffold attachment factor A by heterogeneous nuclear ribonucleoprotein particle-associated PRMT1.* J Biol Chem, 2004. **279**(47): p. 48774-9.
263. Romig, H., et al., *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, 1992. **11**(9): p. 3431-40.
264. Auclair, Y. and S. Richard, *The role of arginine methylation in the DNA damage response.* DNA Repair (Amst), 2013. **12**(7): p. 459-65.
265. Adams, M.M., et al., *53BP1 oligomerization is independent of its methylation by PRMT1.* Cell Cycle, 2005. **4**(12): p. 1854-61.
266. El-Andaloussi, N., et al., *Methylation of DNA polymerase beta by protein arginine methyltransferase 1 regulates its binding to proliferating cell nuclear antigen.* Faseb j, 2007. **21**(1): p. 26-34.
267. Guendel, I., et al., *Methylation of the tumor suppressor protein, BRCA1, influences its transcriptional cofactor function.* PLoS One, 2010. **5**(6): p. e11379.
268. Chen, D.-H., et al., *Effects of Adenosine Dialdehyde Treatment on In Vitro and In Vivo Stable Protein Methylation in HeLa Cells.* The Journal of Biochemistry, 2004. **136**(3): p. 371-376.
269. Cheng, D., et al., *Small molecule regulators of protein arginine methyltransferases.* J Biol Chem, 2004. **279**(23): p. 23892-9.
270. Fujii, S., et al., *Treatment with a Global Methyltransferase Inhibitor Induces the Intranuclear Aggregation of ALS-Linked FUS Mutant In Vitro.* Neurochem Res, 2016. **41**(4): p. 826-35.
271. Polotskaia, A., et al., *Regulation of arginine methylation in endothelial cells: role in premature senescence and apoptosis.* Cell Cycle, 2007. **6**(20): p. 2524-30.
272. Schwerk, C. and K. Schulze-Osthoff, *Methyltransferase inhibition induces p53-dependent apoptosis and a novel form of cell death.* Oncogene, 2005. **24**(47): p. 7002-11.

273. Hong, S., et al., *Methyltransferase-inhibition interferes with neuronal differentiation of P19 embryonal carcinoma cells.* Biochem Biophys Res Commun, 2008. **377**(3): p. 935-40.
274. van der Linden, I.J., et al., *Inhibition of methylation and changes in gene expression in relation to neural tube defects.* Birth Defects Res A Clin Mol Teratol, 2008. **82**(10): p. 676-83.
275. Shiota, M., et al., *Methyltransferase inhibitor adenosine dialdehyde suppresses androgen receptor expression and prostate cancer growth.* J Urol, 2012. **188**(1): p. 300-6.
276. Huang, L., et al., *Arginine methylation of the C-terminus RGG motif promotes TOP3B topoisomerase activity and stress granule localization.* Nucleic Acids Res, 2018. **46**(6): p. 3061-3074.
277. Rhoads, S.N., et al., *The prionlike domain of FUS is multiphosphorylated following DNA damage without altering nuclear localization.* Mol Biol Cell, 2018. **29**(15): p. 1786-1797.
278. Cory, J.G., et al., *Inhibition of nucleic acid synthesis in Ehrlich tumor cells by periodate-oxidized adenosine and adenylic acid.* Arch Biochem Biophys, 1974. **160**(2): p. 495-503.
279. Branscombe, T.L., et al., *PRMT5 (Janus kinase-binding protein 1) catalyzes the formation of symmetric dimethylarginine residues in proteins.* Journal of Biological Chemistry, 2001. **276**(35): p. 32971-32976.
280. Zhou, Z., et al., *PRMT5 regulates Golgi apparatus structure through methylation of the golgin GM130.* Cell Res, 2010. **20**(9): p. 1023-33.
281. Andreu-Perez, P., et al., *Protein arginine methyltransferase 5 regulates ERK1/2 signal transduction amplitude and cell fate through CRAF.* Sci Signal, 2011. **4**(190): p. ra58.
282. Pollack, B.P., et al., *The human homologue of the yeast proteins Skb1 and Hsl7p interacts with Jak kinases and contains protein methyltransferase activity.* J Biol Chem, 1999. **274**(44): p. 31531-42.
283. Mersaoui, S.Y., et al., *Arginine methylation of the DDX5 helicase RGG/RG motif by PRMT5 regulates resolution of RNA:DNA hybrids.* Embo j, 2019. **38**(15): p. e100986.
284. Pal, S., et al., *Human SWI/SNF-associated PRMT5 methylates histone H3 arginine 8 and negatively regulates expression of ST7 and NM23 tumor suppressor genes.* Mol Cell Biol, 2004. **24**(21): p. 9630-45.
285. Wang, L., S. Pal, and S. Sif, *Protein arginine methyltransferase 5 suppresses the transcription of the RB family of tumor suppressors in leukemia and lymphoma cells.* Mol Cell Biol, 2008. **28**(20): p. 6262-77.
286. Deng, X., et al., *Arginine methylation mediated by the Arabidopsis homolog of PRMT5 is essential for proper pre-mRNA splicing.* Proc Natl Acad Sci U S A, 2010. **107**(44): p. 19114-9.
287. Bezzi, M., et al., *Regulation of constitutive and alternative splicing by PRMT5 reveals a role for Mdm4 pre-mRNA in sensing defects in the spliceosomal machinery.* Genes Dev, 2013. **27**(17): p. 1903-16.
288. Jahan, S. and J.R. Davie, *Protein arginine methyltransferases (PRMTs): role in chromatin organization.* Adv Biol Regul, 2015. **57**: p. 173-84.
289. Gayatri, S. and M.T. Bedford, *Readers of histone methylarginine marks.* Biochim Biophys Acta, 2014. **1839**(8): p. 702-10.
290. Duncan, K.W., et al., *Structure and Property Guided Design in the Identification of PRMT5 Tool Compound EPZ015666.* ACS Med Chem Lett, 2016. **7**(2): p. 162-6.
291. Gerhart, S.V., et al., *Activation of the p53-MDM4 regulatory axis defines the anti-tumour response to PRMT5 inhibition through its role in regulating cellular splicing.* Sci Rep, 2018. **8**(1): p. 9711.
292. Scaglione, A., et al., *PRMT5-mediated regulation of developmental myelination.* Nat Commun, 2018. **9**(1): p. 2840.
293. Gostissa, M., et al., *Activation of p53 by conjugation to the ubiquitin-like protein SUMO-1.* Embo j, 1999. **18**(22): p. 6462-71.
294. Rodriguez, M.S., C. Dargemont, and R.T. Hay, *SUMO-1 conjugation in vivo requires both a consensus modification motif and nuclear targeting.* J Biol Chem, 2001. **276**(16): p. 12654-9.

295. Cubenas-Potts, C. and M.J. Matunis, *SUMO: a multifaceted modifier of chromatin structure and function.* Dev Cell, 2013. **24**(1): p. 1-12.
296. Sampson, D.A., M. Wang, and M.J. Matunis, *The small ubiquitin-like modifier-1 (SUMO-1) consensus sequence mediates Ubc9 binding and is essential for SUMO-1 modification.* J Biol Chem, 2001. **276**(24): p. 21664-9.
297. Hendriks, I.A., et al., *Site-specific characterization of endogenous SUMOylation across species and organs.* Nat Commun, 2018. **9**(1): p. 2456.
298. Wei, W., et al., *A stress-dependent SUMO4 sumoylation of its substrate proteins.* Biochem Biophys Res Commun, 2008. **375**(3): p. 454-9.
299. Cappadocia, L., A. Pichler, and C.D. Lima, *Structural basis for catalytic activation by the human ZNF451 SUMO E3 ligase.* Nat Struct Mol Biol, 2015. **22**(12): p. 968-75.
300. Cappadocia, L. and C.D. Lima, *Ubiquitin-like Protein Conjugation: Structures, Chemistry, and Mechanism.* Chem Rev, 2018. **118**(3): p. 889-918.
301. Kunz, K., T. Piller, and S. Muller, *SUMO-specific proteases and isopeptidases of the SENP family at a glance.* J Cell Sci, 2018. **131**(6).
302. Shiio, Y. and R.N. Eisenman, *Histone sumoylation is associated with transcriptional repression.* Proc Natl Acad Sci U S A, 2003. **100**(23): p. 13225-30.
303. Ouyang, J., et al., *Direct binding of CoREST1 to SUMO-2/3 contributes to gene-specific repression by the LSD1/CoREST1/HDAC complex.* Mol Cell, 2009. **34**(2): p. 145-54.
304. Antila, C.J.M., et al., *Sumoylation of Notch1 represses its target gene expression during cell stress.* Cell Death Differ, 2018. **25**(3): p. 600-615.
305. Yu, F., et al., *SUMO suppresses and MYC amplifies transcription globally by regulating CDK9 sumoylation.* Cell Res, 2018. **28**(6): p. 670-685.
306. Nacerddine, K., et al., *The SUMO pathway is essential for nuclear integrity and chromosome segregation in mice.* Dev Cell, 2005. **9**(6): p. 769-79.
307. Chong, P.A., R.M. Vernon, and J.D. Forman-Kay, *RGG/RG Motif Regions in RNA Binding and Phase Separation.* J Mol Biol, 2018. **430**(23): p. 4650-4665.
308. Chen, D.H., et al., *Effects of adenosine dialdehyde treatment on in vitro and in vivo stable protein methylation in HeLa cells.* J Biochem, 2004. **136**(3): p. 371-6.
309. Licatalosi, D.D. and R.B. Darnell, *Applications of Next-Generation Sequencing Rna Processing and Its Regulation: Global Insights into Biological Networks.* Nature Reviews Genetics, 2010. **11**(1): p. 75-87.
310. Kolpa, H.J., F.O. Fackelmayer, and J.B. Lawrence, *SAF-A Requirement in Anchoring XIST RNA to Chromatin Varies in Transformed and Primary Cells.* Dev Cell, 2016. **39**(1): p. 9-10.
311. McHugh, C.A., et al., *The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3.* Nature, 2015. **521**(7551): p. 232-6.
312. Poulter, L., et al., *Locations and immunoreactivities of phosphorylation sites on bovine and porcine tau proteins and a PHF-tau fragment.* J Biol Chem, 1993. **268**(13): p. 9636-44.
313. Redondo, J., et al., *Purkinje Cell Pathology and Loss in Multiple Sclerosis Cerebellum.* Brain Pathol, 2015. **25**(6): p. 692-700.
314. Arao, Y., et al., *A nuclear matrix-associated factor, SAF-B, interacts with specific isoforms of AUF1/hnRNP D.* Arch Biochem Biophys, 2000. **380**(2): p. 228-36.
315. Chi, B., et al., *Interactome analyses revealed that the U1 snRNP machinery overlaps extensively with the RNAP II machinery and contains multiple ALS/SMA-causative proteins.* Sci Rep, 2018. **8**(1): p. 8755.
316. Huttlin, E.L., et al., *Architecture of the human interactome defines protein communities and disease networks.* Nature, 2017. **545**(7655): p. 505-509.
317. Hein, M.Y., et al., *A human interactome in three quantitative dimensions organized by stoichiometries and abundances.* Cell, 2015. **163**(3): p. 712-23.
318. Zhao, Y. and Y.H. Lin, *Whole-cell protein identification using the concept of unique peptides.* Genomics Proteomics Bioinformatics, 2010. **8**(1): p. 33-41.

319. Abdelmohsen, K., et al., *Ubiquitin-mediated proteolysis of HuR by heat shock.* Embo j, 2009. **28**(9): p. 1271-82.
320. Behrends, C., et al., *Network organization of the human autophagy system.* Nature, 2010. **466**(7302): p. 68-76.
321. Li, M., et al., *Lentiviral DDX46 knockdown inhibits growth and induces apoptosis in human colorectal cancer cells.* Gene, 2015. **560**(2): p. 237-44.
322. Li, B., et al., *Knockdown of DDX46 inhibits proliferation and induces apoptosis in esophageal squamous cell carcinoma cells.* Oncol Rep, 2016. **36**(1): p. 223-30.
323. Jiang, F., et al., *Knockdown of DDX46 Inhibits the Invasion and Tumorigenesis in Osteosarcoma Cells.* Oncol Res, 2017. **25**(3): p. 417-425.
324. McCormack, E.A., et al., *Point mutations in a hinge linking the small and large domains of beta-actin result in trapped folding intermediates bound to cytosolic chaperonin CCT.* J Struct Biol, 2001. **135**(2): p. 198-204.
325. Kaisari, S., et al., *Role of CCT chaperonin in the disassembly of mitotic checkpoint complexes.* Proc Natl Acad Sci U S A, 2017. **114**(5): p. 956-961.
326. Balchin, D., M. Hayer-Hartl, and F.U. Hartl, *In vivo aspects of protein folding and quality control.* Science, 2016. **353**(6294): p. aac4354.
327. Gestaut, D., et al., *The ATP-powered gymnastics of TRiC/CCT: an asymmetric protein folding machine with a symmetric origin story.* Curr Opin Struct Biol, 2019. **55**: p. 50-58.
328. Smith, B.N., et al., *Exome-wide rare variant analysis identifies TUBA4A mutations associated with familial ALS.* Neuron, 2014. **84**(2): p. 324-31.
329. Xia, Y., et al., *Pathogenic mutation of UBQLN2 impairs its interaction with UBXD8 and disrupts endoplasmic reticulum-associated protein degradation.* J Neurochem, 2014. **129**(1): p. 99-106.
330. Shtivelman, E., F.E. Cohen, and J.M. Bishop, *A human gene (AHNAK) encoding an unusually large protein with a 1.2-microns polyionic rod structure.* Proc Natl Acad Sci U S A, 1992. **89**(12): p. 5472-6.
331. Shin, J.H., et al., *Increased Cell Proliferations and Neurogenesis in the Hippocampal Dentate Gyrus of Ahnak Deficient Mice.* Neurochem Res, 2015. **40**(7): p. 1457-62.
332. Hutter, K., et al., *SAFB2 Enables the Processing of Suboptimal Stem-Loop Structures in Clustered Primary miRNA Transcripts.* Mol Cell, 2020. **78**(5): p. 876-889.e6.
333. Han, J., et al., *The Drosha-DGCR8 complex in primary microRNA processing.* Genes Dev, 2004. **18**(24): p. 3016-27.
334. Alarcon, C.R., et al., *N6-methyladenosine marks primary microRNAs for processing.* Nature, 2015. **519**(7544): p. 482-5.
335. O'Brien, J., et al., *Overview of MicroRNA Biogenesis, Mechanisms of Actions, and Circulation.* Front Endocrinol (Lausanne), 2018. **9**: p. 402.
336. Nishi, K., et al., *Human TNRC6A is an Argonaute-navigator protein for microRNA-mediated gene silencing in the nucleus.* RNA, 2013. **19**(1): p. 17-35.
337. Gregersen, L.H., et al., *MOV10 Is a 5' to 3' RNA helicase contributing to UPF1 mRNA target degradation by translocation along 3' UTRs.* Mol Cell, 2014. **54**(4): p. 573-85.
338. Antar, L.N., et al., *Localization of FMRP-associated mRNA granules and requirement of microtubules for activity-dependent trafficking in hippocampal neurons.* Genes Brain Behav, 2005. **4**(6): p. 350-9.
339. Kao, D.I., et al., *Altered mRNA transport, docking, and protein translation in neurons lacking fragile X mental retardation protein.* Proc Natl Acad Sci U S A, 2010. **107**(35): p. 15601-6.
340. Zhao, Y., et al., *Activation function-1 domain of androgen receptor contributes to the interaction between subnuclear splicing factor compartment and nuclear receptor compartment. Identification of the p102 U5 small nuclear ribonucleoprotein particle-binding protein as a coactivator for the receptor.* J Biol Chem, 2002. **277**(33): p. 30031-9.

341. Bertram, K., et al., *Cryo-EM Structure of a Pre-catalytic Human Spliceosome Primed for Activation.* Cell, 2017. **170**(4): p. 701-713 e11.
342. Perez, Y., et al., *RSRC1 mutation affects intellect and behaviour through aberrant splicing and transcription, downregulating IGFBP3.* Brain, 2018. **141**(4): p. 961-970.
343. Li, X., et al., *AGO2 and its partners: a silencing complex, a chromatin modulator, and new features.* Crit Rev Biochem Mol Biol, 2020. **55**(1): p. 33-53.
344. Marzec, M., *New insights into the function of mammalian Argonaute2.* PLoS Genet, 2020. **16**(11): p. e1009058.
345. Ashburner, M., et al., *Gene ontology: tool for the unification of biology. The Gene Ontology Consortium.* Nat Genet, 2000. **25**(1): p. 25-9.
346. Huang da, W., B.T. Sherman, and R.A. Lempicki, *Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists.* Nucleic Acids Res, 2009. **37**(1): p. 1-13.
347. Huang da, W., B.T. Sherman, and R.A. Lempicki, *Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources.* Nat Protoc, 2009. **4**(1): p. 44-57.
348. Kuleshov, M.V., et al., *Enrichr: a comprehensive gene set enrichment analysis web server 2016 update.* Nucleic Acids Res, 2016. **44**(W1): p. W90-7.
349. Liao, Y., et al., *WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs.* Nucleic Acids Res, 2019. **47**(W1): p. W199-W205.
350. Szklarczyk, D., et al., *STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets.* Nucleic Acids Res, 2019. **47**(D1): p. D607-D613.
351. Kramer, A., et al., *Causal analysis approaches in Ingenuity Pathway Analysis.* Bioinformatics, 2014. **30**(4): p. 523-30.
352. Oesterreich, S., et al., *Tamoxifen-bound estrogen receptor (ER) strongly interacts with the nuclear matrix protein HET/SAF-B, a novel inhibitor of ER-mediated transactivation.* Mol Endocrinol, 2000. **14**(3): p. 369-81.
353. Pakos-Zebrucka, K., et al., *The integrated stress response.* EMBO Rep, 2016. **17**(10): p. 1374-1395.
354. Hetz, C. and F.R. Papa, *The Unfolded Protein Response and Cell Fate Control.* Mol Cell, 2018. **69**(2): p. 169-181.
355. Liu, C.Y. and R.J. Kaufman, *The unfolded protein response.* J Cell Sci, 2003. **116**(Pt 10): p. 1861-2.
356. Hashimoto, T., et al., *Scaffold attachment factor B: distribution and interaction with ERalpha in the rat brain.* Histochem Cell Biol, 2020. **153**(5): p. 323-338.
357. Hong, E.A., et al., *SAFB1- and SAFB2-mediated transcriptional repression: relevance to cancer.* Biochem Soc Trans, 2012. **40**(4): p. 826-30.
358. Darrow, M.C., et al., *Structural Mechanisms of Mutant Huntingtin Aggregation Suppression by the Synthetic Chaperonin-like CCT5 Complex Explained by Cryoelectron Tomography.* J Biol Chem, 2015. **290**(28): p. 17451-61.
359. Gagnon, K.T., et al., *RNAi factors are present and active in human cell nuclei.* Cell Rep, 2014. **6**(1): p. 211-21.
360. Miao, L., et al., *A dual inhibition: microRNA-552 suppresses both transcription and translation of cytochrome P450 2E1.* Biochim Biophys Acta, 2016. **1859**(4): p. 650-62.
361. Benhamed, M., et al., *Senescence is an endogenous trigger for microRNA-directed transcriptional gene silencing in human cells.* Nature cell biology, 2012. **14**(3): p. 266-275.
362. Pare, J.M., et al., *Hsp90 regulates the function of argonaute 2 and its recruitment to stress granules and P-bodies.* Mol Biol Cell, 2009. **20**(14): p. 3273-84.
363. Patel, R.C. and G.C. Sen, *PACT, a protein activator of the interferon-induced protein kinase, PKR.* EMBO J, 1998. **17**(15): p. 4379-90.

364. Will, C.L. and R. Luhrmann, *Spliceosome structure and function.* Cold Spring Harb Perspect Biol, 2011. **3**(7).
365. Lee, Y.B., et al., *SAFB re-distribution marks steps of the apoptotic process.* Experimental cell research, 2007. **313**(18): p. 3914-23.
366. Leidal, A.M., et al., *The LC3-conjugation machinery specifies the loading of RNA-binding proteins into extracellular vesicles.* Nat Cell Biol, 2020.
367. Mizutani, A., et al., *Boat, an AXH domain protein, suppresses the cytotoxicity of mutant ataxin-1.* EMBO J, 2005. **24**(18): p. 3339-51.
368. Treiber, T., et al., *A Compendium of RNA-Binding Proteins that Regulate MicroRNA Biogenesis.* Mol Cell, 2017. **66**(2): p. 270-284 e13.
369. Rhoads, R.E. and B.J. Lamphear, *Cap-independent translation of heat shock messenger RNAs.* Curr Top Microbiol Immunol, 1995. **203**: p. 131-53.
370. Peters, G.A., et al., *Modular structure of PACT: distinct domains for binding and activating PKR.* Mol Cell Biol, 2001. **21**(6): p. 1908-20.
371. Huang, X., B. Hutchins, and R.C. Patel, *The C-terminal, third conserved motif of the protein activator PACT plays an essential role in the activation of double-stranded-RNA-dependent protein kinase (PKR).* Biochem J, 2002. **366**(Pt 1): p. 175-86.
372. Patel, C.V., et al., *PACT, a stress-modulated cellular activator of interferon-induced double-stranded RNA-activated protein kinase, PKR.* J Biol Chem, 2000. **275**(48): p. 37993-8.
373. Ito, T., M. Yang, and W.S. May, *RAX, a cellular activator for double-stranded RNA-dependent protein kinase during stress signaling.* J Biol Chem, 1999. **274**(22): p. 15427-32.
374. Lee, Y., et al., *The role of PACT in the RNA silencing pathway.* EMBO J, 2006. **25**(3): p. 522-32.
375. Sevier, C.S. and C.A. Kaiser, *Formation and transfer of disulphide bonds in living cells.* Nat Rev Mol Cell Biol, 2002. **3**(11): p. 836-47.
376. Smith, M.H., H.L. Ploegh, and J.S. Weissman, *Road to ruin: targeting proteins for degradation in the endoplasmic reticulum.* Science, 2011. **334**(6059): p. 1086-90.
377. Carrara, M., et al., *Noncanonical binding of BiP ATPase domain to Ire1 and Perk is dissociated by unfolded protein CH1 to initiate ER stress signaling.* Elife, 2015. **4**.
378. Cui, W., et al., *The structure of the PERK kinase domain suggests the mechanism for its activation.* Acta Crystallogr D Biol Crystallogr, 2011. **67**(Pt 5): p. 423-8.
379. Bertolotti, A., et al., *Dynamic interaction of BiP and ER stress transducers in the unfolded-protein response.* Nat Cell Biol, 2000. **2**(6): p. 326-32.
380. Rozpedek-Kaminska, W., et al., *The PERK-Dependent Molecular Mechanisms as a Novel Therapeutic Target for Neurodegenerative Diseases.* Int J Mol Sci, 2020. **21**(6).
381. Nopoulos, P.C., *Huntington disease: a single-gene degenerative disorder of the striatum.* Dialogues Clin Neurosci, 2016. **18**(1): p. 91-8.
382. Ruz, C., et al., *Proteotoxicity and Neurodegenerative Diseases.* Int J Mol Sci, 2020. **21**(16).
383. Dash, D. and T.A. Mestre, *Therapeutic Update on Huntington's Disease: Symptomatic Treatments and Emerging Disease-Modifying Therapies.* Neurotherapeutics, 2020.
384. Orr, H.T., et al., *Expansion of an unstable trinucleotide CAG repeat in spinocerebellar ataxia type 1.* Nat Genet, 1993. **4**(3): p. 221-6.
385. Zoghbi, H.Y. and H.T. Orr, *Pathogenic mechanisms of a polyglutamine-mediated neurodegenerative disease, spinocerebellar ataxia type 1.* J Biol Chem, 2009. **284**(12): p. 7425-9.
386. Tejwani, L. and J. Lim, *Pathogenic mechanisms underlying spinocerebellar ataxia type 1.* Cell Mol Life Sci, 2020.
387. Huo, X., et al., *The Nuclear Matrix Protein SAFB Cooperates with Major Satellite RNAs to Stabilize Heterochromatin Architecture Partially through Phase Separation.* Mol Cell, 2020. **77**(2): p. 368-383.e7.
388. Shevtsov, S.P. and M. Dundr, *Nucleation of nuclear bodies by RNA.* Nat Cell Biol, 2011. **13**(2): p. 167-73.

389. Minguez, P., et al., *Deciphering a global network of functionally associated post-translational modifications.* Mol Syst Biol, 2012. **8**: p. 599.
390. Duan, G. and D. Walther, *The roles of post-translational modifications in the context of protein interaction networks.* PLoS Comput Biol, 2015. **11**(2): p. e1004049.
391. Velazquez-Cruz, A., et al., *Post-translational Control of RNA-Binding Proteins and Disease-Related Dysregulation.* Front Mol Biosci, 2021. **8**: p. 658852.
392. Miller, J.L. and P.A. Grant, *The role of DNA methylation and histone modifications in transcriptional regulation in humans.* Subcell Biochem, 2013. **61**: p. 289-317.
393. Ho, M.C., et al., *Structure of the Arginine Methyltransferase PRMT5-MEP50 Reveals a Mechanism for Substrate Specificity.* Plos One, 2013. **8**(2).