NUCLEOLAR R-LOOPS AS CONTRIBUTORS TO SPINAL MUSCULAR ATROPHY PATHOPHYSIOLOGY

By Nelly Berrueta Ramirez

Submitted for the degree of Doctor of Philosophy (PhD)

Department of Molecular Biology and Biotechnology

School of Biosciences

University of Sheffield

March 2023

Acknowledgements

Life is a rollercoaster or a Jumanji game as I like to refer to my own, you can make plans or avoid certain experiences or behaviours but at the end, the unexpected gets to you and sometimes amazing things that resemble miracles happen. Other times you get entangled in webs of chaos and challenges that force you to show who you REALLY are. For me this time it was the latter. My experience still remains ineffable and have no idea how I survive it. The illness, dread, pain, and trauma unrelated to my studies had an impact on them that was unfortunate, eventually these things got intertwined; and with tears in my eyes, page after page I just hope that the words written in this document be a testament to my love for science and resilience.

I have the need to thank every single person that showed compassion and shared words of kindness during my difficult situation; these represented a lifeline more than you could possibly know. Thank you to my family for showing me their love and actively participate in my therapy and rehabilitation. Thank you to the University of Sheffield, School of Bioscience staff for their advice and help when I needed to take time off to recover. To my mental health adviser Matthew Heaver and thesis mentor Rachael Staves that were by my side (virtually) during my recovery, I do not know what would have become of me without you. Special thanks to my primary supervisor Sherif El-Khamisy for his guidance, and his patience when I was unable to get in touch due to my inability to cope with my condition, thank you so much! My gratitude to my secondary supervisor Mimoun Azzouz for allowing me to be part of this research project; and to Evangelia Karyka for being one of the best laboratory mentors I ever had. To Adelina Acosta Martin for her unwavering friendship and showing me the wonderful world of proteins. Thanks to CONACyT for their financial support which made my studies possible.

I want to show my appreciation for my laboratory years at Firth court and SITraN all the hours of work were always filled with joy and excitement, like playing your favourite games every day. I cherish the friendships, conversations, laughs, with everyone. The PhD was the best learning-holiday ever, relaxing, and brilliant. As we keep living, new

surprises await around the corner, I am healthy now and a different person from who I was before, with a heart and mind open, in search for the next adventure. Things happen FOR us, not TO us. And so, I leave you hoping that you can see your own hardships with love and patience to finally reborn.

P.S. Almost forgot to thank/recommend the ‘Cracking the Cryptic’ YouTube channel that helped me to rehabilitate my brain by generating new neural connections whilst playing the most remarkable sudoku puzzles created by international masters.

Abstract

Spinal Muscular Atrophy (SMA) is an autosomal recessive disorder characterized by degeneration of lower motor neurons, that leads to muscle weakness, and eventually death. SMA is caused by low levels of survival motor neuron (SMN) protein, due to deletion or mutation of SMN1 gene. SMN is a ubiquitously expressed protein, primarily involved in pre-mRNA splicing through its “canonical” function as chaperone in the assembly of small nuclear Ribonucleoproteins (snRNP); additionally, it is believed that it also plays a role during transcription termination by interacting with RNA pol II and Senataxin, a helicase involved in the resolution of DNA/RNA hybrids (R-loops). However, the mechanism through which depletion of SMN disturbs genome stability and contributes to the SMA phenotype remains unclear.

In this project I investigated the presence of R-loops in human cells to determine if these structures contribute to the physiopathology of SMA. R-loops are three stranded nucleic acid structures, usually formed during transcription when the messenger RNA (mRNA) hybridises with the complementary DNA strand, leaving the other DNA strand exposed. For this experiment, fibroblasts from SMA patients overexpressing an enhanced green fluorescent protein (EGFP), fused with the hybrid binding domain (HB) from RNaseH1, an endonuclease that resolves R-loops was used. The fusion protein acted as a R-loop capture agent for the Chromatin immunoprecipitation-quantitative polymerase chain reaction (ChIP-qPCR) approach. Further, I analysed the levels of other markers and proteins associated with DNA damage, to assess the degree of cellular stress present in SMA.

XX REDACTED TEXT XX

Table of Contents

List of Figures

List of Tables

Abbreviations

A-T: Ataxia Telangiectasia

A: Adenine

AAV: Adeno-associated viral vector

AB: Ammonium Bicarbonate

ACN: Acetonitrile

ALS: Amyotrophic Lateral Sclerosis

AMP: *Adenosine* monophosphate

AP: abasic site

APS: Ammonium persulfate

ASO: Antisense oligonucleotide

ATM: Ataxia Telangiectasia Mutated

ATM: ataxia telangiectasia mutated

ATR: Ataxia Telangiectasia and Rad3 related

BER: Base Excision Repair

REDACTED

REDACTED

BisMapR: bisulfite MapR

BLM: Bloom syndrome

BRCA1: breast cancer 1

BRCA2: breast cancer 2

BSA: Bovine Serum Albumin

C: Cytosine

ChIP: Chromatin Immunoprecipitation

CpG: Cytosine-Guanine dinucleotide

CPT: Camptothecin

CRISPR: clustered regularly interspaced short palindromic repeats

CSA: Cockayne syndrome A

CSB: Cockayne syndrome B

CTD: carboxyl-terminus domain

CtIP: C-terminal binding protein interacting protein

CUT&Tag:

ddH2O

DDR: DNA damage response

DDX: DEAD-box

DNA Pol: DNA polymerase

DNA-PK: DNA-dependent Protein Kinase

DNA: Deoxyribonucleic Acid

DRIP: DNA/RNA Immunoprecipitation

DSB: Double Strand Break

dsRNA: double-stranded RNA

DTT: dithiothreitol

E. coli: Escherichia coli

ECL: Electrochemiluminescence

EDTA: *Ethylenediaminetetraacetic acid*

EV: empty vector

FA: Fanconi Anaemia

FA: Formic acid

FAAP24: Fanconi Anaemia associated protein 24

FANCM: Fanconi anaemia complementation group M

FBS: Foetal bovine serum

FCS: Foetal calf serum

G: Guanine

GFP: Green Fluorescent Protein

GG-NER: global genome-NER

GndHCl: Guanidine – Hydrochloric acid

HB: hybrid binding

HBD: hybrid binding domain

HBSS: EDTA- Hanks' balanced salt solution

HEK: Human embryonic kidney

HeLa: Henrietta Lacks

HEPES: N-(2-Hydroxyethyl) piperazine-N'-(2-ethanesulfonic acid)

HR: Homologous Recombination

HRE: hexanucleotide repeat expansion

HRP: horseradish peroxidase

IAA: iodoacetamide

ICL: interstrand crosslinks

IF: Immunofluorescence

IP: immunoprecipitation

iPSc: induced pluripotent stem cells

IR: ionizing radiation

ITS: Internal transcribed spacer

KD: Knock-down

kDa: kilo Daltons

LIG1: DNA ligase 1

LIG3: DNA ligase 3

lncRNA: long noncoding RNA

MN: Motor neuron

MNase: micrococcal nuclease

MNPs: motor neuron progenitors

MOI: multiplicity of infection

mRNA: messenger RNA

mRNP: messenger ribonucleprotein

MS: Mass spectrometry

MSH: MutS Homolog

MW: molecular weight

NB: no biotin

ncRNA: non-coding RNA

NHEJ: Non-homologous End Joining

NT: non transfected

OAO2: Oculomotor Apraxia Type 2

PALB2: partner and localiser of BRCA2

PARP: poly-ADP Ribose Polymerase

PARPi: PARP inhibitor

PBS: Phosphate-buffered saline

PCNA: replication proteins proliferating cell nuclear antigen

PEI: polyethylenimine

PFA: Paraformaldehyde

PLB: protein loading buffer

POL β: DNA polymerase β

POL δ/ε: DNA polymerase δ/ε

qPCR: quantitative Polymerase Chain Reaction

QS: quantum satis (enough amount)

R-loop: RNA/DNA hybrid

rDNA: ribosomal DNA

RIPA: Radioimmunoprecipitation Assay

RNA Pol: RNA Polymerase

RNA: Ribonucleic Acid

RNase H: Ribonuclease H

ROS: Reactive oxygen species

RPA: replication protein A

rRNA: ribosomal RNA

RT: room temperature

S9.6: anti-R-loop antibody

SCAN1: Spinocerebellar ataxia with axonal neuropathy

SDS: sodium dodecyl sulphate

Seq: sequencing

SETX: senataxin

SFM: Serum-free media

shRNA: small hairpin RNA

sincRNAs: sense intergenic noncoding RNAs

siRNA: small interfering RNA

siSMA: small interfering SMA

SITraN: Sheffield institute for Translational Neuroscience

SMA: Spinal Muscular Atrophy

SMN: survival of motor neuron

snRNP: small nuclear Ribonucleoprotein

SSB: Single Strand Break

ssDNA: single strand DNA

T: Thymine

TBE: Tris Borate EDTA

TBS: Tris Buffer Saline

TBST: Tris Buffer Saline Tween

TC-NER: transcription coupled NER

TCEP: tris (2-carboxyethyl) phosphine

TFIIH: transcription initiation factor II H

TOP: Topoisomerase

TOP1cc: Topoisomerase 1 covalent complex

U: Uracil

UAF1: USP1-associated factor

USP1: ubiquitin carboxy-terminal hydrolase I

UV-DDB: UV-damaged DNA binding protein

UV: ultraviolet

W/V: weight over volume

XPA: xeroderma pigmentosum A

XPB: xeroderma pigmentosum B

XPC: xeroderma pigmentosum C

XPD: xeroderma pigmentosum D

XRCC1: X-ray repair cross-complementing protein 1

γH2AX: phosphorylated histone 2

1. CHAPTER I. INTRODUCTION

1.1. DNA DAMAGE REPAIR

1.1.1. DNA damage

Deoxyribonucleic acid (DNA) is the blueprint that harbours all the information for the development and cellular functions of living beings. Thus, the preservation of DNA integrity is vital for the overall wellbeing of organisms. The genome is constantly under the attack of several factors that can damage the DNA (Figure 1.1). Each human cell experiences approximately over 10,000 DNA lesions per day, which can be a result of regular cellular processes or external agents. Consequently, robust DNA repair mechanisms have been developed by cells to protect DNA and maintain genome integrity. Accumulation of DNA damage in cells can lead to senescence, cancer, and neurodegenerative disorders (Chatterjee and Walker, 2017; Carusillo and Mussolino, 2020; Yousefzadeh et al., 2021).

1.1.1.1. Types of DNA damage

According to its origin, DNA damage can be classified into two categories: endogenous and exogenous. Endogenous damage refers primarily to lesions generated from within the cells, that can result from interactions between the DNA and reactive molecules. This type of damage is usually continuous and persistent. Conversely, exogenous DNA damage is caused by external sources such as environmental factors, chemical or physical agents, which can be very potent (Chatterjee and Walker, 2017; Yousefzadeh et al., 2021).

1.1.1.1.1. Endogenous DNA damage

Endogenous DNA damage include diverse types of lesions, and each can have different consequences. For example, replication errors, can generate base mismatches; here, deoxynucleotides can be inserted, deleted, or substituted during DNA synthesis and cause mutations. During base deamination, a nitrogenous base loses its exocyclic amine; this is considered a major source of mutagenesis. Next, abasic or apurinic/apyrimidic

(AP) sites can be generated by cleavage or unprompted hydrolysis of the N-glycosyl bond between the sugar phosphate and the nitrogenous base. These abasic sites are unstable and can easily turn into single strand breaks (SSB) after the cleavage of the 3' phosphodiester bond from the AP site by beta-elimination. Oxidative DNA damage is caused by reactive oxygen species (ROS) and can have deleterious consequences. DNA methylation can be highly mutagenic or generate AP sites. Finally, mutations or genome instability can also be the result of perturbances in the DNA repair pathways or caused by cellular stress when biological processes surpass the capacity of the cell to repair the DNA damage (Chatterjee and Walker, 2017; Tubbs and Nussenzweig, 2017).

DNA lesion	Apurinic site	Strand cross-linking	Bulky Lesion	Mismatch	Double-Stranded Break
Main causes	Cellular metabolism	UV radiation	UV radiation	Replication errors	IR; Replication fork stallation
DDR	BER	FANCONI	NER	MMR	NHEJ; MMEJ; SSA; HDR

Figure 1.1. Graphic representation of major DNA lesions. Table below denotes the type of DNA injury, the primary source of damage and the DNA repair pathway involved in its resolution (From Carusillo and Mussolino, 2020).

1.1.1.1.2. Exogenous DNA damage

Among the exogenous DNA damage sources radiation can have a strong negative impact on DNA. Ionizing radiation (IR), which is constituted by neutrons, alpha, beta, gamma, and x-rays. Such radiation can damage the DNA in a similar way as ROS. It can also cause SSBs and double strand breaks (DSB). Ultraviolet (UV) radiation, the primary cause of human skin cancer, is classified into three types, according to their wavelength range. UV-C (190-290 nm) is the most hazardous type since the maximum wavelength of UV

absorption for DNA is 260 nm. This radiation mainly generates covalent bonds between neighbouring pyrimidines, creating bulky lesions that distort the DNA helix. UV-B (290-320 nm) can also generate pyrimidine dimers, albeit less efficiently than UV-C. Finally, UV-A (320-400 nm) induces photooxidation reactions that can cause adduct formation over the DNA; this can ultimately result in DNA strand breaks. Chemical exogenous sources that can cause DNA damage include alkylating agents, such as smoke, tobacco, or chemotherapeutic molecules. Aromatic amines from fuel, coal, pesticides, transform into carcinogenic alkylating agents in the body. Polycyclic aromatic hydrocarbon are carcinogenic compounds that result from combustion and are widely spread in the environment. Finally other sources like toxins, extreme temperatures or hypoxia have also been shown to generate DNA damage (Rastogi et al., 2010; Kantidze et al., 2016; Chatterjee and Walker, 2017).

1.1.2. DNA Repair

DNA lesions are detected by a system of sensor proteins that constantly monitor the integrity of the genome to promote DNA repair. This is called the DNA damage response (DDR). These DDR factors are recruited to the site according to the type of damage in order to activate the appropriate DNA repair pathway (Chatterjee and Walker, 2017; Carusillo and Mussolino, 2020).

1.1.2.1. DNA repair pathways

The maintenance of genome integrity is of the utmost importance to ensure survival; therefore, the repair mechanisms that protect DNA are robust and specialised in different types of lesions. There are five major DNA repair pathways, which are active throughout the cell cycle: base excision repair, nucleotide excision repair, mismatch repair, homologous recombination and non-homologous end joining (Chatterjee and Walker, 2017; Carusillo and Mussolino, 2020).

1.1.2.1.1. Base excision repair (BER)

BER is specialised in the repair of non-bulky single base lesions such as oxidation, alkylation, or deamination; this repair process mostly activates during the G1 phase of the cell cycle. The injuries are firstly recognised by DNA glycosylases, which removes the aberrant base by cleaving the N-glycosylic bond between the base and the sugar generating an AP site (Figure 1.2). Within the 11 glycosylases that participate in BER, some are monofunctional (only glycosylase activity) or can be bifunctional (with a β-lyase in addition to the glycosylase activity). This difference in function determines the progression towards the short patch of BER when the AP site is generated by monofunctional glycosylases, and towards the long patch BER in the case of bifunctional glycosylases. Once the AP site is generated, PARP-1 and PARP-2 bind to it and recruit the human AP endonuclease (APE1), which nicks the DNA backbone 5' to the AP site. In the short patch repair, DNA polymerase β (POLβ) fills the single nucleotide gap; and DNA is sealed by DNA ligase 1 (LIG1), or by a complex of X-ray repair cross-complementing protein 1 (XRCC1) and DNA ligase 3 (LIG3). Finally, in the long patch repair, the nucleotide gap is filled by POL β, or DNA polymerase δ/ε (POL δ/ε); followed by the removal of the flap, and the DNA is sealed by LIG1 (Chatterjee and Walker, 2017; Lee and Kang, 2019; Carusillo and Mussolino, 2020).

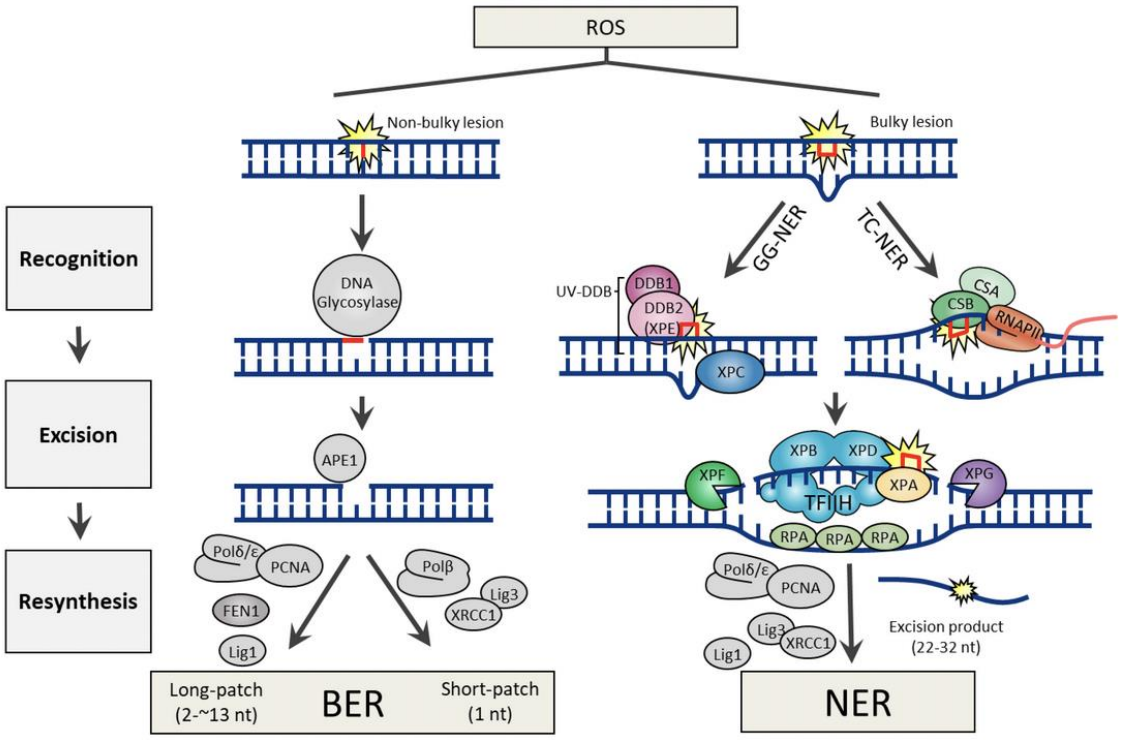

Figure 1.2. Diagram of BER and NER repair mechanisms. Image displays DNA excision repair mechanisms for bulky or non-bulky lesions generated by ROS (From Lee and Kang, 2019).

1.1.2.1.2. Nucleotide excision repair (NER)

Nucleotide excision repair pathway successfully resolves a broad range of DNA lesions, including interstrand crosslinks and bulky adducts (Figure 1.2). Additionally, is the only pathway that repairs bulky photolesions caused by exposure to UV radiation. NER has two main branches for repair: global genome-NER (GG-NER), and transcription coupled NER (TC-NER). In GG-NER the bulky lesions are recognised by the xeroderma pigmentosum C (XPC) protein, or the UV-damaged DNA binding protein (UV-DDB). On the other hand, when the lesion occurs over a transcribed gene, this is considered as an RNA polymerase II blockage. Then, the damage is recognised by the Cockayne syndrome B (CSB) and Cockayne syndrome A (CSA). The following steps are conserved on both branches. The transcription initiation factor II H (TFIIH) is recruited, from which xeroderma pigmentosum group A (XPA) confirms presence of damage; the helicases XPB and XPD (xeroderma pigmentosum group B and C) unwind the DNA, and the replication protein A (RPA) stabilises the unharmed DNA strand. The damaged DNA fragment is removed by the endonucleases XPF and XPG (xeroderma pigmentosum group F and G). Finally, the filling of the gap and ligation are carried out by the replication proteins proliferating cell nuclear antigen (PCNA), POL δ, POL ε and LIG1 or by the complex XRCC1-LIG3 (Chatterjee and Walker, 2017; Lee and Kang, 2019; Carusillo and Mussolino, 2020).

1.1.2.1.3. Mismatch repair (MMR).

Mismatch repair is a post replicative DNA repair pathway that primarily resolves mismatched base pairs that are usually generated by replication errors. These DNA lesions are firstly sensed by MutS Homolog (MSH) complexes (Figure 1.3). The MutSα heterodimer (MSH2/MSH6) identifies mismatches, and one-to-two nucleotide insertion-deletion loops (IDL). Whereas the MutSβ heterodimer (MSH2/MSH3) identifies longer IDLs. After damage recognition MutS shifts conformationally to form a

stable bond with the mismatched region; later, it recruits MutL to the damage site. MutLα heterodimer (MLH1/PMS2) behaves as a mediator of the strand excision, and later, the exonuclease 1 (EXO1) removes the nucleotides surrounding the DNA lesion (Chatterjee and Walker, 2017; Carusillo and Mussolino, 2020; Pećina-Šlaus et al., 2020).

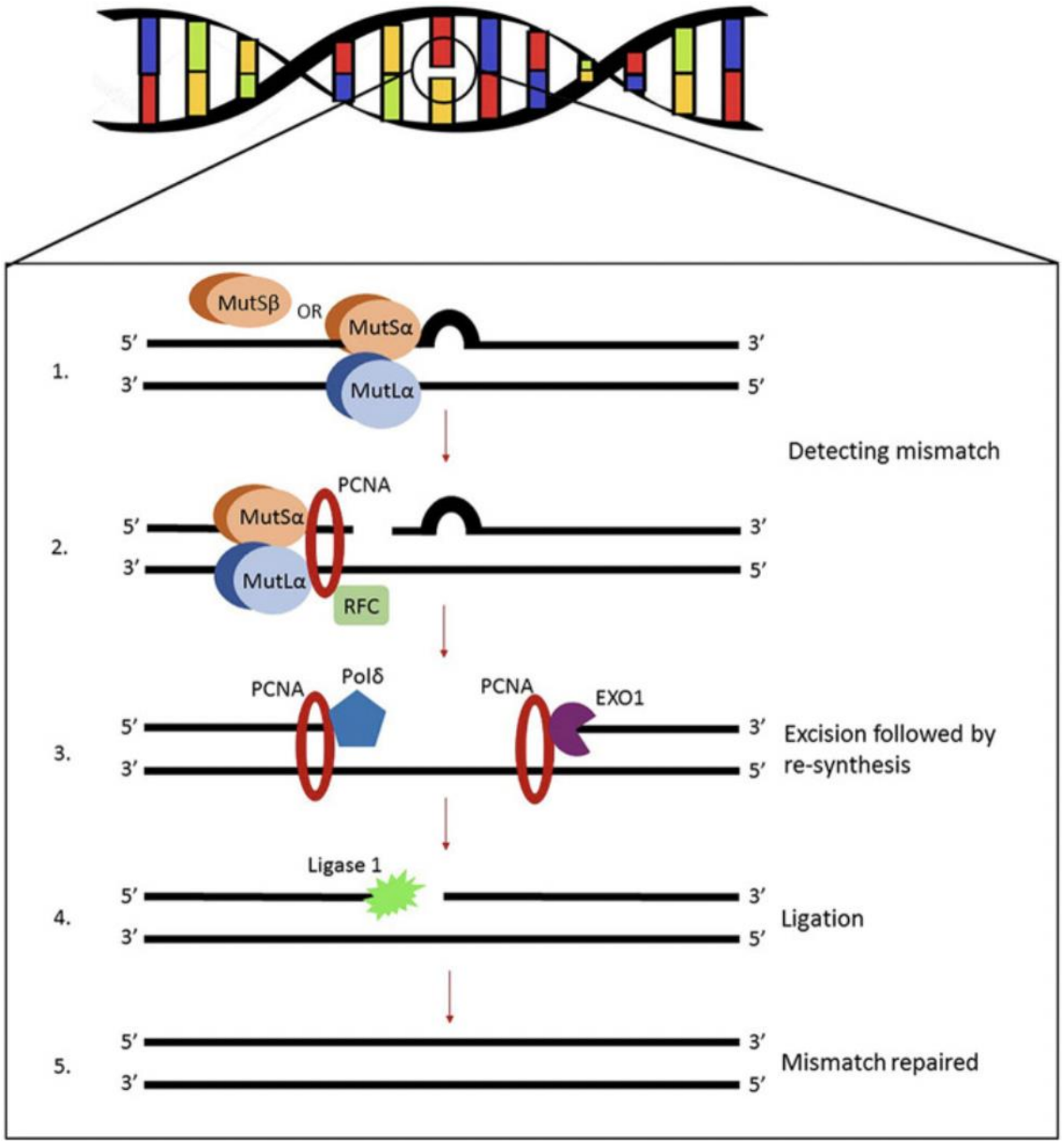

Figure 1.3. Diagram of the mismatch repair mechanism. Image represents DNA repair through the mismatch mechanisms that begins by the recognition of the base pair mismatches in the helix. (From Pećina-Šlaus et al., 2020)

1.1.2.1.4. Fanconi anaemia repair (FA).

The Fanconi anaemia repair pathway, also known as the Fanconi anaemia/BRCA pathway, includes 21 complementation groups (A-V) within its protein network. It resolves interstrand crosslinks (ICL) that form between complementary DNA strands as a result of the activity of crosslinking agents. Moreover, repair of ICLs is triggered by the convergence of replication forks into the crosslinks; and deficiencies in their repair can cause DNA breaks and cancer predisposition. Upon ICL presence in the genome, UHRF1 and Fanconi anaemia complementation group M (FANCM) sense the damage and bind to the ICL site along with FAAP24 (Fanconi Anaemia associated protein 24), MFH1 (also known as FAAP1) and MFH2 (also known as FAAP10). FANCM then promotes ATR response which recruits the proteins of the Fanconi anaemia core complex to the damaged site and activates them by phosphorylation (Figure 1.4). Next, the FA core complex ubiquitin ligase subunit FANCL interacts with the E2 enzyme UBE2T to monoubiquitylate the FANCD2-FANCI heterodimer. Meanwhile, breast cancer protein1 (BRCA1) displaces the replicative helicase CMG complex from the DNA strand, possibly through its ubiquitin ligase activity. This permits one replication fork to 'approach' to the ICL within one nucleotide apart, and the nearing of ubiquitylated FANCD2 into the ICL region as well. The next step is to create a DNA incision on each side of the crosslinked nucleotide. Here, ubiquitylated FANCD2 recruits the nucleases SLX4 and FAN1; additionally, SLX4 recruits endonucleases such as ERCC4 (also known as XPF, which is also involved in NER pathway), MUS8-EME1, among others, resulting in the 'unhooking' of the DNA strands with the crosslinked nucleotide attached to the complementary strand (blue section form Figure 1.4). Later, translesion synthesis polymerases (REV1 or POLζ) proceed to DNA synthesis bypassing the lesion; and ligation restores the first DNA duplex (peach section in Figure 1.4), which subsequently is used as a template for homologous recombination repair of the DSB created by the endonuclease cleaving (green section of Figure 1.4). Finally, deubiquitylation of FANCD2 and FANC1 by the ubiquitin carboxy-terminal hydrolase I (USP1) and USP1-associated factor (UAF1), releases the proteins from the DNA and finishes repair. The remaining bulky lesion can be repaired by NER (Ceccaldi et al., 2016; Chatterjee and Walker, 2017; Gueiderikh et al., 2022).

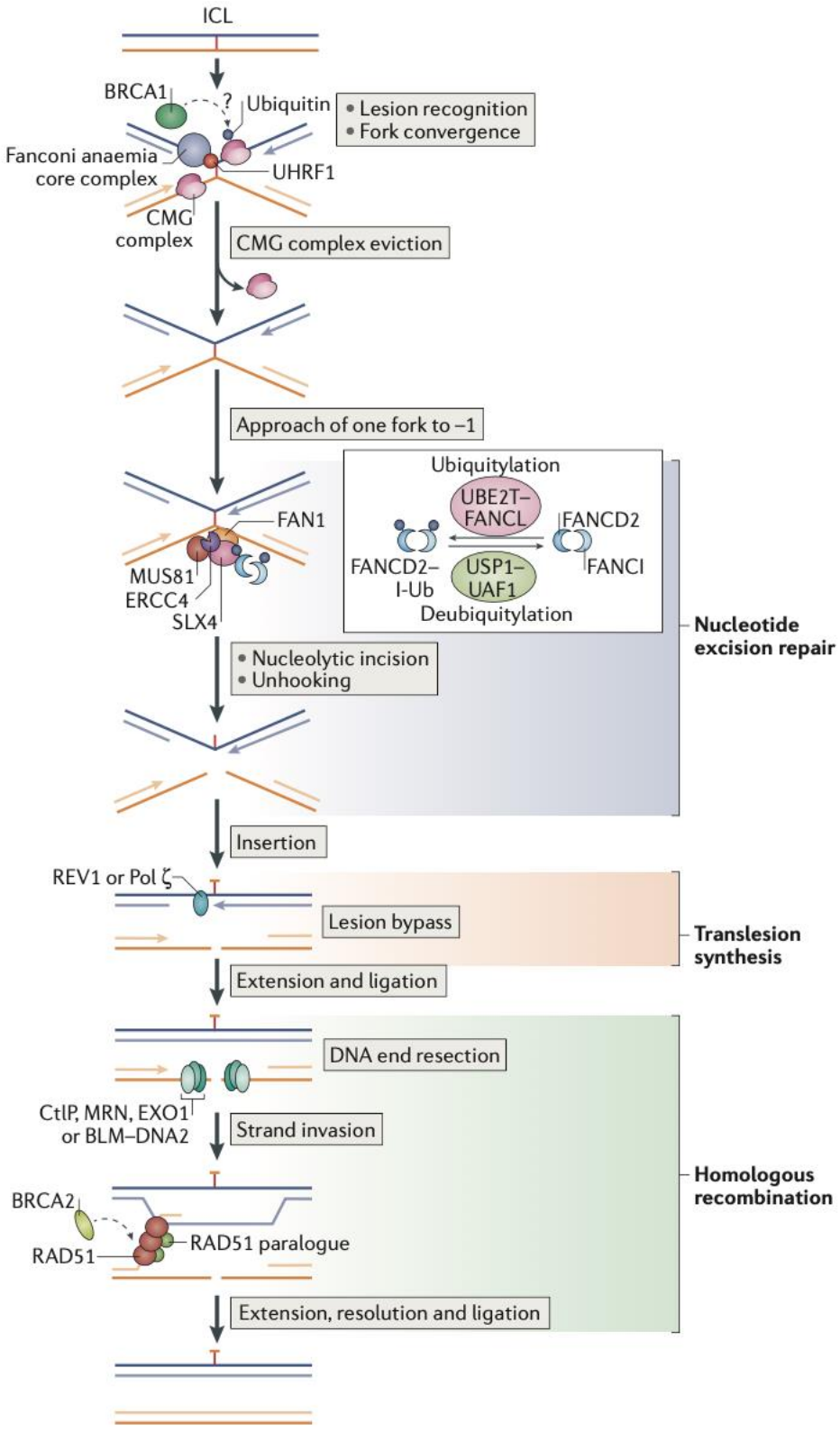

Figure 1.4. Graphic representation of the Fanconi Anaemia repair pathway. Image shows an overview of DNA repair through the FA pathway, where nucleotide excision repair (NER) and homologous recombination (HR) cooperate to resolve interstrand crosslinks (ICL). (From Ceccaldi et al., 2016)

1.1.2.1.5. Homologous recombination (HR)

The homologous recombination repair pathway (Figure 1.5) relies in the DNA homology of a sister chromatid to resolve double strand breaks (DSB) with high-fidelity. The broken DNA strands are first recognised by the MRN (MRE11-RAD50- NBS1) complex, which binds to both damaged ends. Next, MRE11 starts resection and NBS1 recruits the ataxia telangiectasia mutated (ATM) protein, which phosphorylates H2AX to serve as a platform for other repair factors. This is followed by short nucleotide resection by the exonuclease C-terminal binding protein interacting protein (CtIP). Later, a long-range resection is performed by EXO1, the endonuclease DNA2 and the Bloom syndrome (BLM) helicase. The single strand DNA (ssDNA) that resulted from the resection process is then coated by the RPA complex to preserve the integrity of the DNA. This recruits the ataxia telangiectasia and Rad3-related protein (ATR) and ATRIP that phosphorylate the kinase Chk1. After, RPA coating is replaced by Rad51 that forms nucleoprotein filaments with the DNA. For this, BRCA2 is recruited to the DSB site by partner and localiser of BRCA2 (PALB2), that later form a complex with BRCA1-BARD1. Later the nucleoprotein filament generated invades the homologous region of its sister chromatid, forming a D-loop where the DNA synthesis carried out by POL δ, κ and ν (Chatterjee and Walker, 2017; Carusillo and Mussolino, 2020; Sun et al., 2020).

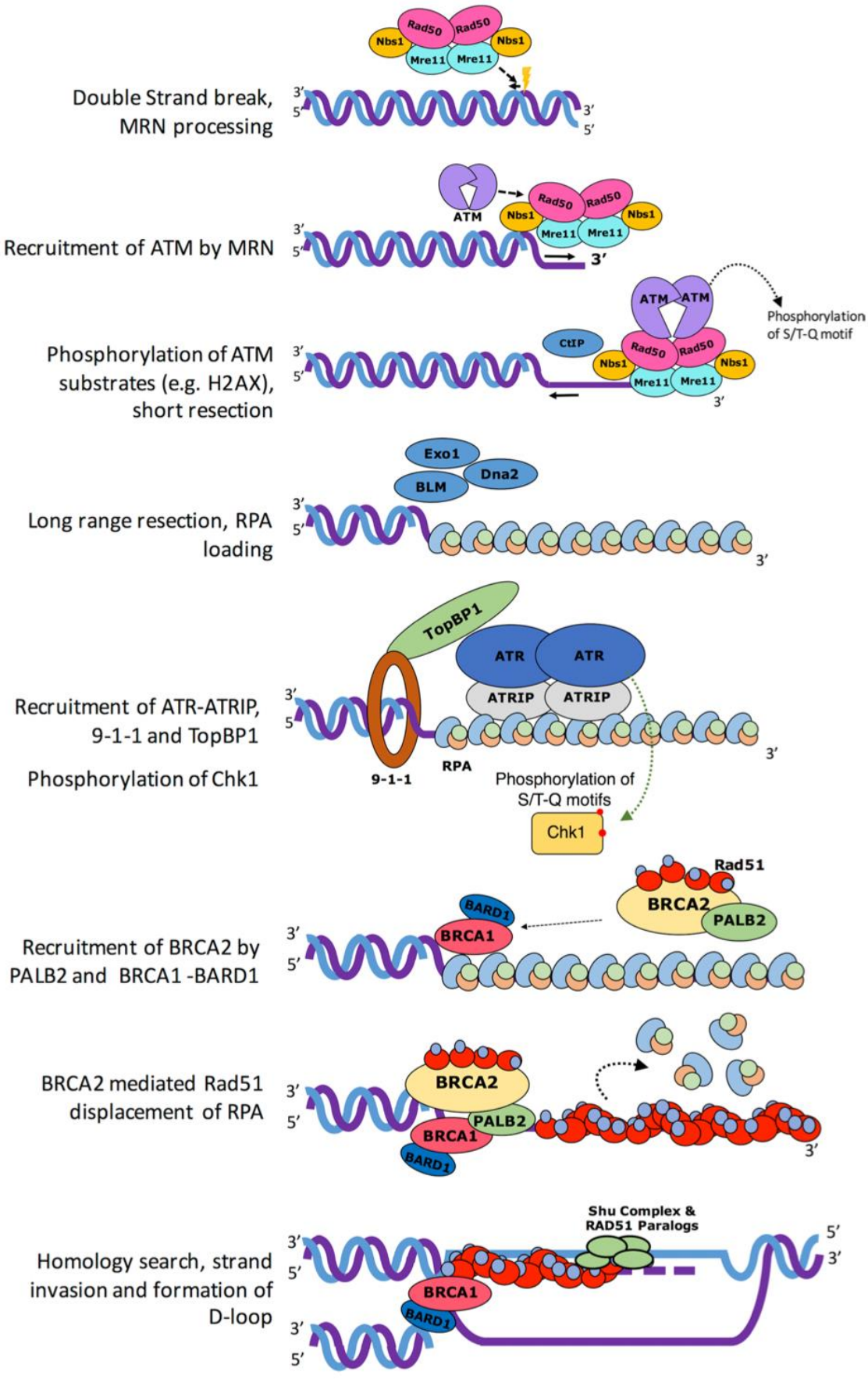
Double Strand break, MRN processing
Rad50
Rad50
Nbs1
Nbs1
Mre11
Mre11
3′
5′
Recruitment of ATM by MRN
ATM
Phosphorylation of ATM substrates (e.g. H2AX), short resection
Phosphorylation of S/T-Q motif
CtIP
Long range resection, RPA loading
Exo1
Dna2
BLM
Recruitment of ATR-ATRIP, 9-1-1 and TopBP1
Phosphorylation of Chk1
TopBP1
ATR
ATRIP
9-1-1
RPA
Phosphorylation of S/T-Q motifs
Chk1
Recruitment of BRCA2 by PALB2 and BRCA1 -BARD1
Rad51
BRCA2
PALB2
BRCA1
BARD1
BRCA2 mediated Rad51 displacement of RPA
Homology search, strand invasion and formation of D-loop
Shu Complex & RAD51 Paralogs

Figure 1.5. Representation of the DNA damage response signalling in Homologous recombination. Image displays a summary of the protein signaling during the first stages of the homologous recombination repair pathway (From Sun et al., 2020).

1.1.2.1.6. Non-homologous end joining (NHEJ)

The non-homologous end joining pathway (Figure 1.6) is the main mechanism for the resolution of DSB in dividing and non-dividing somatic cells. However, this pathway oftentimes gives rise to small deletions and insertions due to the nature of DNA end processing. NHEJ requires little to no microhomology (less than 20 bp) among the DNA overhangs. The first element to identify the damaged site is the Ku (Ku70 and Ku 80) heterodimer. This heterodimer protects the DNA ends from resection and recruits the DNA-dependent protein kinase (DNA-PK). Later, DNA-PK phosphorylates nearby components of the NHEJ pathway and recruits the X-ray repair cross complementing protein 4 (XRCC4) and DNA ligase4 (LIG4) that stabilise the NHEJ complex. Next, Ku70-Ku80 and XRCC4-LIG4 mediate synapsis and direct ligation for both DNA blunt ends and overhangs. Additionally, the XRCC4-like factor (XLF) and PAXX (paralogue of XRCC4 and XLF) can also be recruited to the complex to promote synapsis and ligation (peach section in Figure 1.6). On the other hand, when DNA ends are not suitable for direct ligation through XRCC4-LIG4, they can be first processed by POL μ and POL λ to fill in nucleotides and mediate synapsis, or by the nuclease Artemis and tyrosyl-DNA phosphodiesterase 1 (TDP1), to remove nucleotides from incompatible DNA ends; thus, rendering them suitable for ligation (purple section in Figure 1.6). Finally, depending on the nature of the damage on DNA ends and their processing, small additions and deletions can be encountered at the repaired junctions. (Chatterjee and Walker, 2017; Carusillo and Mussolino, 2020; Zhao et al., 2020).

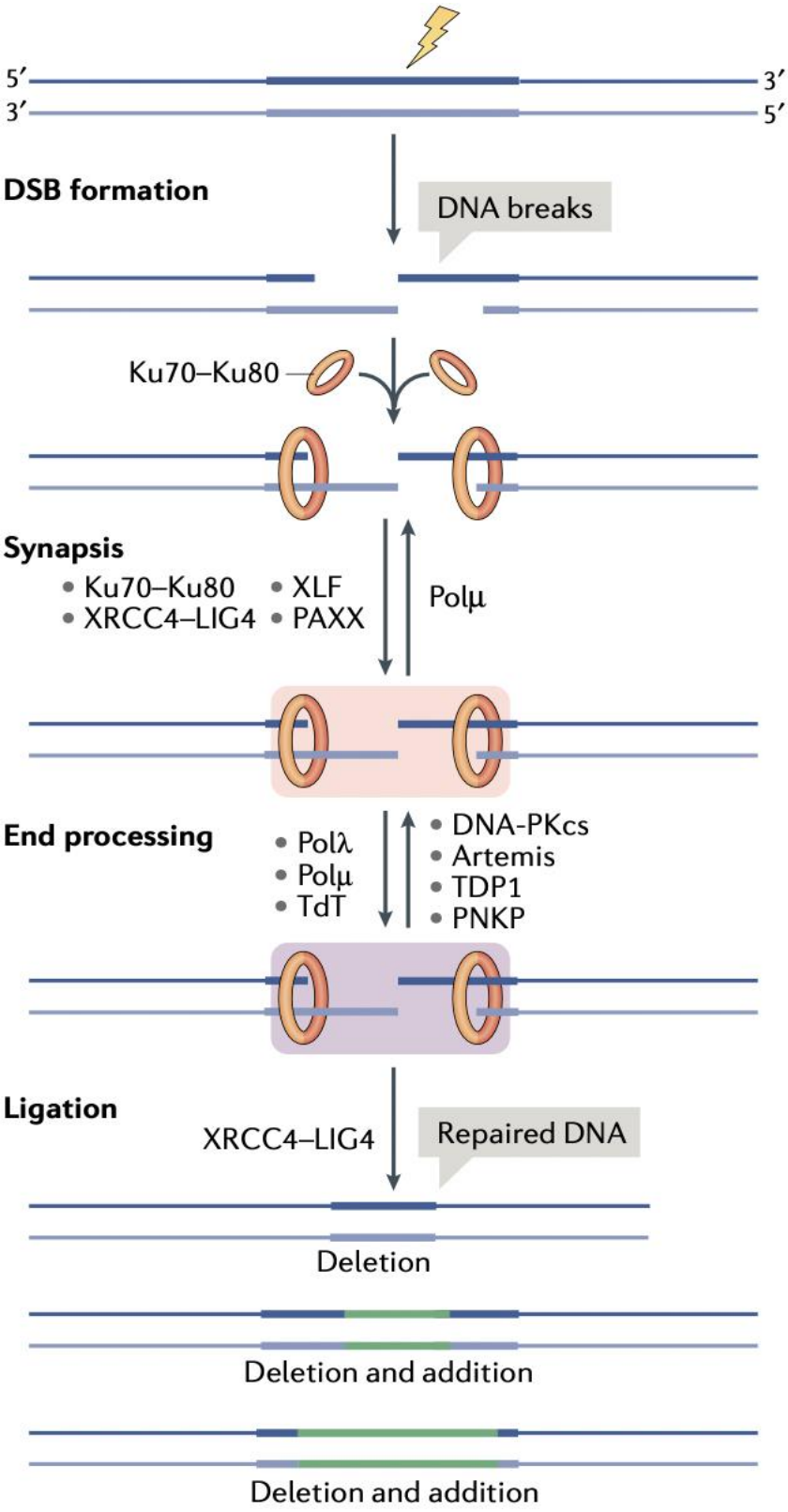

Figure 1.6. Non-homologous end joining repair pathway. Graphic representation of the non-homologous end joining repair process following a double strand break (From Zhao et al., 2020).

1.2. TRANSCRIPTION AND DNA REPAIR

1.2.1. Transcription-associated DNA damage

DNA, the self-replicating material that in collaboration with environmental influences produces a phenotype, can be damaged physically or chemically by endogenous and exogenous sources. This creates DNA lesions that can be mended by highly specialized mechanisms to restore genome integrity. Chronic exposure to DNA damage can result into genetic information loss and mutations that can pose a potential threat to human health. (Sebastian and Oberdoerffer, 2017; Helena et al., 2018). Among the most persistent sources of endogenous DNA damage, essential cellular process like replication and transcription, have the potential to alter the DNA structure. In this sense, transcription can have greater adverse consequences for all cell types since replication is limited to dividing cells. (Sebastian and Oberdoerffer, 2017; Helena et al., 2018).

Transcription has been linked to DNA damage through recombination, mutagenesis, and formation of DSB. During transcription initiation, topoisomerases relieve torsional stress from the DNA promoter regions. (Ju et al., 2006; Kim and Jinks-Robertson 2012; Sebastian and Oberdoerffer, 2017). These proteins create a transient nick on the DNA to resolve topological issues and facilitate transcription; however, topoisomerase misfunction can prevent re-ligation and generate DSB. Moreover, promoter associated DSBs have been found to contribute to chromosomal translocations in cancer (Ju et al., 2006; Lin et al., 2009; Katyal et al., 2014; Crewe and Madabhushi, 2021).

Transcription can also cause replication stress and genome instability when transcription factors collide head-to-head or co-directionally with the replication machinery over common fragile sites in the genome. These collisions can stall replication forks and generate recombination events as well as DSB across the gene body. Furthermore, DNA/RNA hybrids (R-loops) can arise at sites of transcription-replication collision or during transcription pause events and result in DSB. (Helmrich et al., 2011; Sebastian and Oberdoerffer, 2017).

1.2.2. Transcriptional R-loops

R-loops are three-stranded nucleic acid structures formed by the hybridization of RNA with complementary DNA and a displaced DNA single strand (Figure 1.7). Here, the exposed ssDNA from the R-loop is more labile towards damaging agents including cellular mutagenic enzymes such as cytidine deaminases. This enzyme mediates hydrolytic deamination of deoxycytidine into deoxyuridine, which can result in an abasic site lesion after the subsequent excision of the uracil base by DNA glycosylase (Stirling and Hieter, 2016). Regulation of R-loop structures in cells is of great importance; they are required for biological processes, such as mitochondrial DNA replication, immunoglobulin class switch recombination and regulation of gene expression. Nonetheless, dysregulation of R-loops can promote DNA damage and genome instability in the form of chromosome fragility and hyper-recombination which may lead to human diseases (Groh and Gromak, 2014).

R-loop structures were firstly visualised in vitro in 1976 through electron microscopy (EM) (Thomas et al., 1976) and the hybrids were thermodynamically more stable than double stranded DNA (Groh and Gromak, 2014). Later, the presence of R-loops in living organisms was reported in bacteria in 1994 (Drolet et al., 1994); since, R-loops have been also found in yeast and mammalian cells (Sollier and Cimprich, 2015; Castillo Guzman and Chédin, 2021).

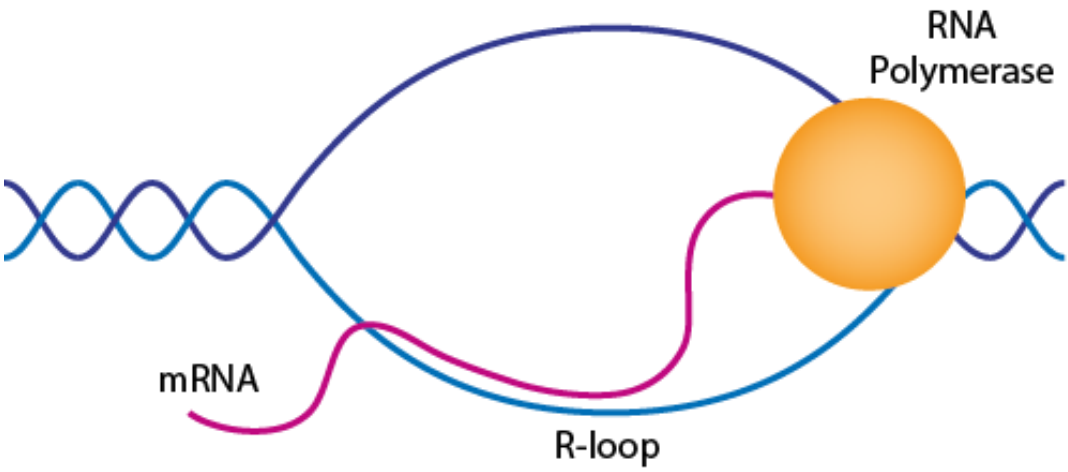

Figure 1.7. Diagram of the structure of an R-loop. During transcription nascent mRNA hybridizes with the template strand and displaces the coding strand, leaving it exposed to damage.

R-loops have been studied through diverse methods. *In vivo* detections include EM, immunofluorescence (IF) and Chromatin immunoprecipitation (ChIP) using a catalytically inactive RNase H1 or a fusion protein containing the hybrid binding (HB) domain from RNase H1 and green fluorescent protein (GFP), (Bhatia et al., 2014; Chen et al., 2017; García-Muse and Aguilera, 2019). RNase H1 is part of a family of endonucleases found in prokaryotes and eukaryotes as well as in reverse transcriptase from retroviruses. This enzyme hydrolyses the phosphodiester bond of the RNA strand in a DNA/RNA hybrid (Hyjek et al., 2019). On the other hand, ex vivo methods mainly rely on the use of S9.6 antibody for DNA/RNA immunoprecipitation (DRIP) or IF (Castillo Guzman and Chédin, 2021). The S9.6 monoclonal antibody was originally obtained by immunization of mice using a synthetic DNA/RNA antigen derived from the bacteriophage ϕX174 that infects *Escherichia* (Phillips et al., 2013). Techniques using this antibody require previous extraction of nucleic acids from cells and do not provide information about R-loops in a more natural environment (Castillo Guzman and Chédin, 2021).

The results from these two detection methods have revealed important differences in R-loop distribution according to the mapping methodology. Recent R-loop profiling comparing an artificial DNA/RNA hybrid sensor that contained the hybrid binding domain from RNase H1 (GST-His_6-2xHBD) against S9.6 antibody, using native and *ex vivo* approaches, showed that the differences in data appear to fluctuate according to the method and not the sensor used (Wang et al., 2021). Native mapping techniques seem to detect best R-loops that form close to promoter regions, and *ex vivo* techniques reflect better R-loops formed over the gene body. Therefore, some scientists have proposed that R-loops can be classified into promoter-associated and elongation-associated hybrids (Castillo Guzman and Chédin, 2021).

R-loops can be generated co-transcriptionally by hybridization of nascent mRNA with the DNA template, or with complementary sequences far from the original transcription site (Sollier and Cimprich, 2015). Also, they can arise as a result from defects in pre-mRNA splicing and during RNA export (Chen et al., 2017). The half-life of DNA/RNA

hybrids is approximately 10 minutes, and their reported length varies from less than 100 bp up to 2kb. Promoter-associated R-loops are shorter with 60 bp or less compared to 300 bp or more from the elongation-associated R-loops (Sanz et al., 2016; García-Muse and Aguilera, 2019; Castillo Guzman and Chédin, 2021).

R-loop formation is increased over regions where they perform physiological functions such as replication of mitochondrial DNA, telomere homeostasis, immunoglobulin class switch recombination, transcription initiation and termination. In general, R-loops accumulate in highly transcribed genes, including the ribosomal, centromeres and telomeres. Molecularly, their occurrence is favoured by the presence of GC-rich and GC-skewed regions in the DNA located over promoter or transcription termination sites, as well as by the formation of G-quadruplexes in the displaced DNA single strand, DNA nicks and negative DNA supercoiling. Further, GC-rich retained introns can also operate as substrates for DNA/RNA hybrid generation (Bhatia et al., 2017; Jangi et al., 2017; García-Muse and Aguilera, 2019).

1.2.2.1. Physiological R-loops

R-loops play crucial modulating roles in several organisms; in immunoglobulin class switch recombination, R-loops are generated during transcription in B cells, to favour recombination (Yu et al., 2003; Groh and Gromak, 2014). DNA/RNA hybrids also behave as intermediaries during initiation of mitochondrial DNA replication; and in bacteria replication, by opening the DNA at the origins of ColE1-type plasmids. (Castillo Guzman and Chédin, 2021). R-loops mediate telomere elongation by promoting recombination in yeast (Balk et al., 2013). During gene editing via CRISPR-Cas9 activity, an R-loop is formed before DNA cleavage by Cas9 (Jinek et al., 2012) (Figure 1.8). Moreover, R-loops also modify the status of the chromatin structure, maintaining it ‘open’ to favour epigenetic regulation (Powell et al., 2013).

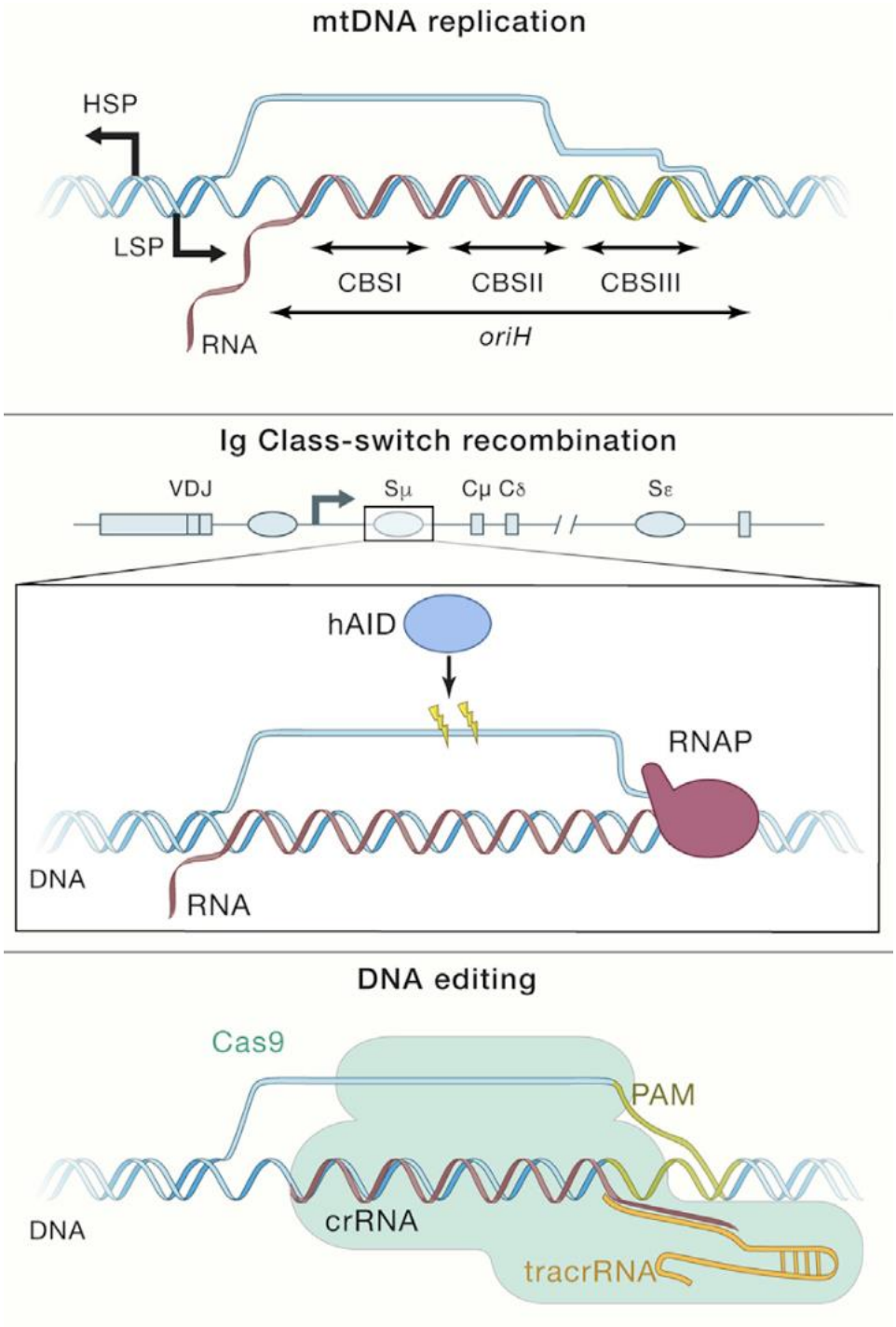

Figure 1.8. Physiological R-loops. Graphic representation of R-loops involved in physiological processes. From top to bottom: R-loop formed over replication origin of mitochondrial DNA; R-loops as intermediaries of Ig Class-switch recombination; and R-loops formed by a guide RNA during CRISPR-Cas9 editing (From García-Muse and Aguilera, 2019).

In transcriptional processes, R-loops have been associated with active gene expression by blocking methylation over more than 1,200 gene promoters (Grunseich et al., 2018). Moreover, they can support antisense transcription by behaving as promoters to generate long noncoding RNAs (lncRNA) (Tan-Wong et al., 2019); or they can block the lncRNA promoters preventing their transcription as observed in *Arabidopsis* (Sun et al., 2013). Furthermore, these lncRNAs can also influence gene expression by binding to gene promoters and forming an R-loop to facilitate demethylation and transcription (Arab et al., 2019). During transcription termination R-loops assist the pausing of RNA polymerase II in collaboration with senataxin (SETX) a helicase that resolves R-loops (Cohen et al., 2018), and survival motor neuron (SMN) protein. SMN recognizes and binds to the symmetric dimethylated arginine residue R1810 (R1810me2s), from the RNA polymerase II Carboxy-terminal domain (CTD), interacts with SETX and stabilizes the interaction between SETX and the CTD domain from RNA polymerase II (Figure 1.10) (Zhao et al., 2016). Further, R-loops enriched over G-rich transcription termination regions, can promote antisense transcription to generate dsRNA and recruit RNA interference (RNAi) factors and G9a histone lysine methyltransferase, to support RNA polymerase II pausing and transcription termination (Skourti-Stathaki et al., 2014).

1.2.2.2. R-loop regulation.

Factors involved in R-loop homeostasis include a wide range of proteins with diverse functions in line with the various roles that R-loops play. These factors include transcription associated, splicing, polyadenylation, export proteins, among others; and primarily prevent the formation and excessive accumulation of R-loops (Groh and Gromak, 2014).

In the prevention of R-loop formation, one mechanism that occurs is the coating of nascent mRNA with pre-mRNA splicing factors; thus, preventing hybridisation with complementary DNA (Li and Manley, 2005; Chakraborty et al., 2018). Further, the RNA-binding and export complex THO/TREX, has been proposed to participate in the inhibition of R-loop formation, not only for its contribution to the assembly of optimal

messenger ribonucleprotein (mRNP) molecule; but also, through the interaction of THO with the histone deacetylase complex Sin3A; hence, promoting chromatin closing (Salas-Armenteros et al., 2017) (Figure 1.9). This is important since chromatin status also plays a role in R-loop formation, and open chromatin is often associated with R-loop peaks (Chen et al., 2017). Moreover, negative supercoiling that may be generated during transcription DNA favours R-loop production; this is regulated by topoisomerases that cleave the DNA and relaxes it, DNA returning it to its original state (Santos-Pereira and Aguilera, 2015).

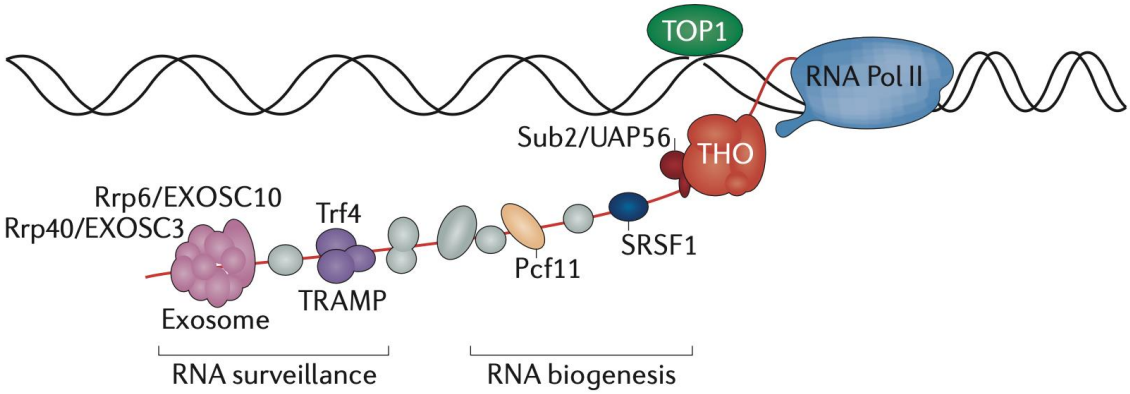

Figure 1.9. R-loop prevention. Diagram of proteins linked to R-loop prevention. Topoisomerase 1 (TOP1) resolves the negative supercoiling; THO complex and other factors involved in RNA biogenesis and surveillance maintain the transcribed mRNA in optimal conditions and coated to prevent hybridization (From Santos-Pereira and Aguilera, 2015).

Additionally, in telomeric regions are comprised by multiple (TTAGGG) repeats coated by 'shelterins' and telomerase complexes to prevent its degradation. The lncRNA telomeric repeat-containing RNA (TERRA) is transcribed by RNA polymerase II over telomeric regions and interacts with telomeric DNA to form R-loops that in excess, can cause replication stress and genome instability. BRCA1 prevents this by binding to TERRA promoters and block its transcription (Toubiana and Selig, 2018; Vohhodina et al., 2021).

To avert R-loop accumulation, the removal of hybridised RNA strand is required, which is carried out by endonucleases and helicases (Santos-Pereira and Aguilera, 2015). In humans, RNase H enzymes hydrolyse the phosphodiester bond of RNA residues in a

metal-dependent manner by binding divalent metal ions like Mg^{2+} or Mn^{2+} (Hyjek et al., 2019). Cleavage by RNAse H1 demands a minimum of four ribonucleotides and the nick site occurs in the middle of the RNA fragment. Conversely RNase H2 cleaves preferably on the 5' side of the RNA stretch and has the faculty to remove individual ribonucleotides (Hyjek et al., 2019).

DNA/RNA helicases facilitate the removal of R-loops by unwinding the RNA strand from the DNA helix in an ATP-dependent manner (Garcia-Muse and Aguilera, 2019). Senataxin (SETX), a 302 kDa protein that localises in the nucleus and cytoplasm, is mostly associated with efficient transcription termination by unwinding DNA/RNA hybrids formed at transcription termination sites (TTS), where it interacts with BRCA1 and SMN (Skourti-Stathaki et al., 2011; Hatchi et al., 2015; Zhao et al., 2016). Further, SETX and SMN are downstream recruiting targets of the zinc finger protein (ZPR1); this protein is required for SETX binding to R-loops (Figure 1.10), and its deficiency cause their accumulation (Kannan et al., 2022). Moreover, SETX proteostasis is mediated by the ubiquitin specific peptidase 11 (USP11) and KEAP1, an E3 ubiquitin ligase; loss of USP11 leads to R-loop accumulation (Jurga et al., 2021).

The Aquarius (AQR) DNA/RNA helicase has been linked to R-loop dependent DNA damage upon absence of the protein (Sollier et al., 2014). Helicase DHX9, was shown to be required for transcription termination and to prevent R-loop formation *in vivo* (Cristini et al., 2018, Zhao et al., 2016). Interestingly, DHX9 has also been suggested to facilitate R-loop formation through the unwinding of secondary RNA structures in cells with defective splicing functions (Chakraborty et al., 2018). DEAD-box helicase 19 (DDX19), has been found to resolve R-loops *in vitro* generated by replication and transcription machinery collision in the presence of DNA damage (Hodroj et al., 2017). The bloom syndrome protein (BLM) is a helicase that is required for the resolution of G-quadruplexes and R-loops (Tan et al., 2020). Other helicases, such as DDX5, DDX17, DDX21, DDX39B, DDX41 have been also linked to R-loop resolution (Song et al., 2017; Mersaoui et al., 2019, Cargill et al., 2021; Boleslavska et al., 2022).

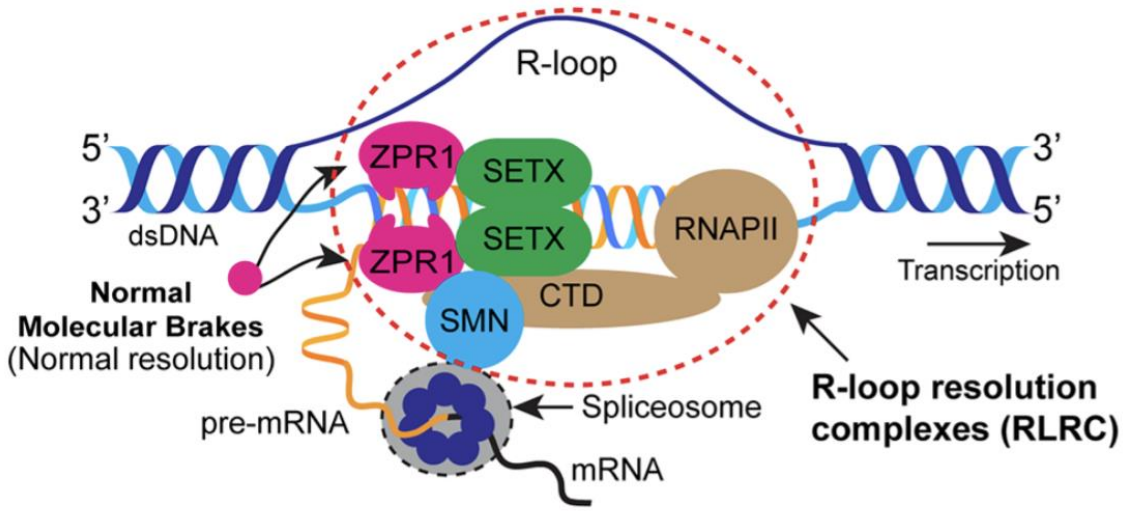

Figure 1.10. R-loop resolution. Proteins associated with resolution of R-loops formed over transcription termination sites, during normal conditions. R-loop resolution complexes formed by sentaxin (SETX), survival motor neuron (SMN), zinc finger protein (ZPR1) and thr C-terminal domain of RNA polymerase II. ZPR is proposed to behave as a molecular brake for R-loop resolution by tethering to the R-loop and recruiting SETX; thus, regulating its activity (From Cuartas and Gangwani 2022).

Finally, unscheduled R-loops can represent a roadblock in the progression of replication forks and result in genome instability; therefore, several factors involved in DNA repair have been associated to R-loop resolution. Depletion of some of the Fanconi Anaemia factors including FANCD2, FANCA, BRCA2, among others, has been shown to produce accumulation of R-loops (Bhatia et al., 2014; García-Rubio et al., 2015; García-Muse and Aguilera, 2019). Further, FANCD2 contributes to R-loop resolution through the recruitment of RNA processing factors hnRNPU and DDX47 (Okamoto et al., 2019); and complexes formed by BRCA1 and SETX are recruited to transcription termination sites to resolve R-loops (Hatchi et al., 2015). Additionally, R-loops can be processed by endonucleases part of the transcription-coupled NER DNA repair pathway, XPF, XPG and FEN1. These factors, cleave the DNA on both extremities of the R-loop, which can induce the formation of double strand breaks (Cristini et al., 2019).

1.3. R-LOOPS AND THEIR ASSOCIATION WITH DISEASE

1.3.1. R-loop dysregulation

Defects in R-loop metabolism caused by mutations or deficiency of factors involved in these processes can generate persistent R-loops that can result in DNA damage and genome instability (Richard and Manley, 2017). Firstly, damage can be triggered in mitotic and post-mitotic cells through injuries in the ssDNA of the R-loop that is left exposed to harming factors. An example of this is the cytidine deaminase (AID), an enzyme that belongs to the APOBEC (apolipoprotein B mRNA editing enzyme, catalytic polypeptide) family, attacks ssDNA converting dC into dU residues that can result in DNA breaks. Moreover, other members of this family have been proposed to have a similar activity in cells that lack AID expression (Sollier and Cimprich, 2015).

Secondly, biological processes can also be hindered by R-loop dysregulation. High levels of R-loops have been correlated with transcription pausing at promotors (Chen et al., 2017). Further, defects in R-loop formation and resolution could also impact transcription since these structures assist the pausing of RNA polymerase II during transcription termination (Cohen et al., 2018). The aforementioned scenarios could be possible in both dividing and non-dividing cells; conversely, in mitotic cells alone an additional source of DNA damage associated to these hybrids takes place. The accumulation of R-loops and paused RNA polymerases can block replication progression, which can cause breakage of the replication fork (García-Muse and Aguilera, 2019).

Senescence, cancers, and neurodegenerative conditions have been associated to R-loops. Loss of R-loop regulators like RNase H or Thp2, a subunit of the THO complex in *S. cerevisiae*, leads to accumulation of telomeric R-loops formed by the lncRNA TERRA in yeast, causing replication stress and promoting senescence. Further, R-loops also mediate telomere elongation through recombination. Yeast deficient in telomerase and Rad52 accumulate R-loops, resulting in telomere shortening and senescence (Balk et al., 2013; Pfeiffer et al., 2013; Groh and Gromak, 2014).

Some of the links between R-loops and cancer are the common fragile sites (CSF); these regions which are prone to R-loop formation and DSB, tend to be rearranged in cancerous cells (Helmrich et al., 2011). Further, *BRCA1* and *BRCA2* genes associated with breast cancer, have been shown to aid in the process of R-loop removal, and mutations in their DNA sequence disrupt their protein function (Richard and Manley, 2016). Additionally, oncogenic events can increase the levels of R-loops and genome instability. The chimeric protein EWS/FLI1 (Ewing's sarcoma/Friend leukaemia integration 1) formed by a translocation, behaves as an aberrant transcription factor, stimulating RNA polymerase II activity and inducing R-loops (May et al., 1993; Petermann et al., 2022). Moreover, oncoproteins such as SS18-SSX1 and HRAS among others, also promote R-loop formation (Petermann et al., 2022). Conversely, some oncogenes can supress the generation of R-loops like *MYCN* and *MDM2* (Klusmann et al., 2018; Herold et al., 2019).

In post-mitotic cells, the excessive presence of R-loops can contribute to the pathophysiology of neurodegenerative diseases. Repetitive trinucleotide expansions like GAA on and CGG on FXN and FMR1 genes associated with Friedreich Ataxia and Fragile X syndrome respectively, induce R-loops that promote silencing of these genes (Figure 1.11) (Groh at al., 2014; Loomis et al., 2014). Moreover, CAG trinucleotide expansions in the *ATXN2* gene encodes for a long polyglutamine (polyQ) tract in the protein. These expansions are associated to ataxia oculomotor apraxia (ALS) and spinocerebellar ataxia type 2 (SCA2). Further, the yeast ortholog of ATXN2 inhibits R-loop formation over G-quadruplexes, and possibly the human protein could fulfil the same role (van Blitterswijk et al., 2014; Perego et al., 2018). TAR DNA-binding protein 43 (TDP43) aggregates, one of the hallmarks of frontotemporal dementia and ALS, impair the self-assembly and nucleic acid-binding function of TDP43, which in turn hinders the R-loop regulation activity of the protein (Wood et al., 2020). Finally, mutations in the genes that encode for SETX helicase, RNAse H enzymes and survival motor neuron (SMN) protein, alter R-loop levels and are associated to ALS4, Aicardi-Goutières Syndrome and Spinal Muscular Atrophy (SMA) respectively (Groh and Gromak, 2014; Cuartas and Gangwani, 2022). Some of these neurological disorders are briefly described below.

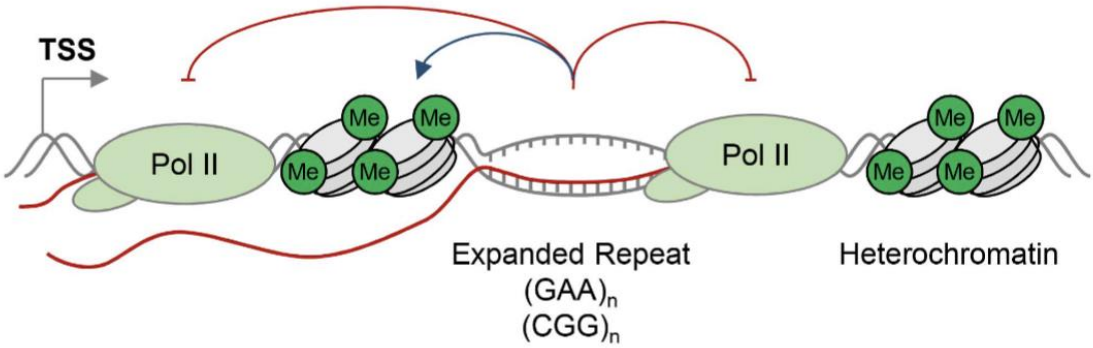

Figure 1.11. Trinucleotide expansion disorders. R-loops formed over trinucleotide repeat expansions block RNA polymerase transcription and favours formation of chromatin marks associated to gene silencing. (From Groh and Gromak, 2014).

1.3.2. Amyotrophic Lateral Sclerosis (ALS)

ALS is characterised by the degeneration of upper motor neurons in the brain and lower motor neurons in the spinal cord; this results in loss of muscle strength, that progresses into paralysis and respiratory failure (Perego et al., 2018). The most common genetic cause of familial and sporadic ALS is an hexanucleotide repeat expansion (HRE) GGGGCC in intron one of the *C9ORF72* (*C9*) gene (Majounie et al., 2012). The association of the HRE with R-loops is not clear yet; however, it has been proposed that perhaps G-rich expansions in the first intron of *C9* might form G-quadruplexes and R-loops; thus, interfering with transcription progression (Perego et al., 2018). Nonetheless, R-loops, double strand breaks and faulty ATM-mediated repair have been identified as part of the pathophysiology of this neurodegenerative condition (Walker et al., 2017).

ALS type 4, an autosomal dominant juvenile form of ALS has been associated with a gain-of-function mutation (L389S) of the helicase SETX; which diminishes formation of R-loops in patients (Figure 1.12) (Groh and Gromak, 2014; García-Muse and Aguilera). Further, it has been proposed that the low levels of R-loops would not be able to prevent DNA methylation at promoters; hence, causing a decrease in transcription of the bone morphogenic protein (BMP) and BAMBI (activin membrane-bound inhibitor), that

modulates negatively the TGF-β (transforming growth factor-β) signalling pathway (Cuartas and Gangwani, 2022).

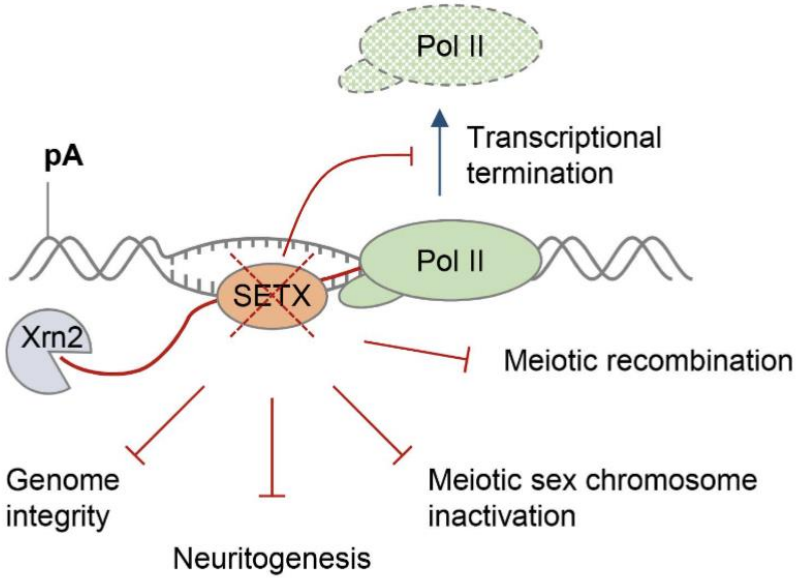

Figure 1.12.R-loop regulation in ALS4 and AOA2. Mutations in Senataxin (SETX) protein alter its helicase capacity to resolve R-loops. This results in deficiencies in several cellular processes like the preservation of genome integrity and neuronal differentiation (From Groh and Gromak, 2014).

1.3.3. Ataxia with oculomotor apraxia type 2 (AOA2)

Ataxia with oculomotor apraxia type 2 (AOA2) is a recessive disorder with an adolescence onset with symptoms that include cerebellar atrophy, oculomotor apraxia, distal muscle weakness and atrophy. AOA2 has been linked with pathogenic variants of SETX caused by loss-of-function mutations (Figure 1.12) (Moreira and Koenig, 2004). The helicase SETX interacts with the Rrp45 exosome subunit, co-localising to R-loop regions. However, this interaction, which is mediated by sumoylation, is altered in AOA2 (Richard and Manley, 2014). This may affect the ability of SETX to resolve R-loops in collaboration with the RNA nuclear exosome. Further, telomere length of AOA2 cells is reduced, which could be justified by the compromised role of SETX to resolve telomeric R-loops (Perego et al., 2018).

1.3.4. Aicardi-Goutières syndrome (AGS)

Aicardi-Goutières Syndrome (AGS) is an inherited autoinflammatory disorder, generated by an excess in production of interferon. Patients present neurologic impairment reflected on muscle weakness of limbs, systemic inflammatory symptoms, and acute intellectual disability, among others (Cristini et al., 2022). AGS has been linked to mutations on RNase H2, an endonuclease that can hydrolyse single ribonucleotides and participates in DNA/RNA hybrid resolution (Figure 1.13). This protein has a trimer structure (A, B, C) with subunit H2C in the middle; although the H2A subunit possesses the catalytic function, all three subunits are necessary for the degradation of RNA. RNase H2 depletion triggers R-loops accumulation and hinders transcription in cells (Hyjek et al., 2019; Cristini et al., 2022). Other genes associated to this condition are part of the nucleic acid sensing machinery that participate in the metabolism of DNA/RNA hybrids, which include, the exonuclease *TREX-1*, dNTP triphosphatase *SMAHD1*, *IFIH1* and dsRNA editing enzyme *ADAR* (Adang et al., 2020).

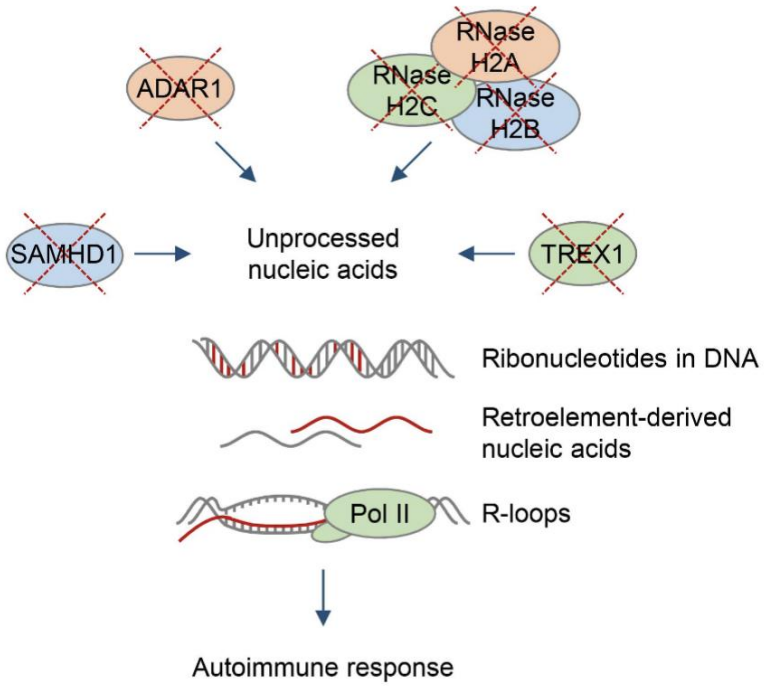

Figure 1.13. Aicardi-Goutières syndrome. Associated with mutations in the three subunits of RNAse H2, a protein that removes ribonucleotides; and factors involved in the processing of nucleic acids like SMAHD1, ADAR1 and TREX1. The unprocessed ribonucleotides in the DNA lead to the autoimmune response observed in AGS (From Groh and Gromak, 2014).

1.4. SPINAL MUSCULAR ATROPHY

Spinal muscular atrophies (SMAs) are hereditary neuromuscular disorders characterized by degeneration of lower motor neurons leading to skeletal muscle weakness, atrophy, and eventually respiratory failure and death (Ahmad et al., 2016; Farrar et al., 2015). Although, more than 30 genes with common pathophysiological pathways have been identified to be associated with these disorders, SMA (Spinal Muscular Atrophy) usually refers to the most common form, which associates with insufficient levels of Survival Motor Neuron protein (SMN) expressed in motor neurons, caused by mutations or homozygous deletion of the Survival Motor Neuron 1 (*SMN1*) gene (Ahmad et al., 2016; Farrar et al., 2015; Lanfranco et al., 2017).

The following most common form is Spinal muscular atrophy with respiratory distress (SMARD1) which on the other hand, is caused by mutations in the *IGHMBP2* (immunoglobulin μ DNA-binding protein) gene. This autosomal recessive disease presents with motor neuron degeneration, muscular atrophy, and diaphragmatic palsy (Perego et al., 2018). In this document, I will continue with the description of the SMN-associated spinal muscular atrophy, which was the subject of this study.

1.4.1. Clinical features of SMA

SMA is the primary monogenic cause of infant mortality, with a prevalence of 1 in 6,000 to 1 in 10,000 worldwide, and a carrier frequency of 1 in 40 to 1 in 60 (Ahmad et al., 2016; Jangi et al., 2017; Scheffer et al., 2001). It was firstly described in the 19th century by Guido Werdnig with features including muscular weakness and the degeneration of anterior horn cells upon autopsy (Mercuri, 2021). The progressive loss of α-motor neurons from the anterior horns of the spinal cord in SMA, translates into weakness and subsequent atrophy of proximal muscles, normally in a symmetrical fashion, causing greater distress on legs than arms. Furthermore, diaphragm, facial and extraocular muscles are relatively less affected; however, children usually require assisted ventilation due to respiratory insufficiency (Farrar et al., 2015; Jangi et al., 2017). Other complications in SMA patients may include increase in contractures; weight gain;

scoliosis, with an 80% prevalence; gastrointestinal dysmotility and cardiac malformations (Nicolau et al; 2021).

SMA presents with a wide variation in clinical progression; thus, it has been typically classified into four types. The categorisation of SMA is based on severity, age of onset and milestones achieved (Table 1.1). Further, the severity of the symptoms in SMA patients has been mainly linked to the levels of SMN protein produced by the survival motor neuron 2 (*SMN2*) gene. *SMN2* is a copy of the *SMN1* gene that holds more than 99% nucleotide identity towards it; however, the 1% difference results in the production of a full-length functional protein in only 10% of its transcripts. The number of *SMN2* copies among individuals has been observed to fluctuate from zero to eight, hence producing different amounts of protein. Lower levels of SMN result in more severe phenotypes, and complete absence of functional SMN protein is lethal. However, the copy number of *SMN2* gene is not the absolute determinant for severity of SMA in all cases, since patients diagnosed with different SMA types can have the same *SMN2* copy number (Scheffer et al., 2001; Calucho et al., 2018; Keinath et al., 2021).

The most common form of SMA is type I (also known as Werdnig-Hoffman disease) which is the most severe. It has an onset before 6 months old and a life expectancy of 2 years. In this scenario infants never accomplish sitting without aid, presenting proximal limb weakness, deficient feeding, and respiratory insufficiency. Additional features include chest deformation into a bell-shape, tongue fasciculations and normal cognitive functions. The majority of type I patients have one to two copies of the *SMN2* gene; nevertheless, three copies have been also found in this category. Type II is intermediate with children showing symptoms before 18 months old. Affected exhibit proximal muscular weakness which is more severe in the lower limbs; tongue fasciculations and atrophy; dysphagia, respiratory insufficiency, and significant scoliosis. They are able to sit by themselves but need assistance to walk for the rest of their lives; commonly reach adulthood. Patients usually present 3 copies of *SMN2*; nonetheless, according to Calucho et al., two and four copies have been found on some occasions (16% and 5%

respectively) (Scheffer et al., 2001; Farrar et al., 2015; Calucho et al., 2018; Schorling et al., 2020; Keinath et al., 2021).

Type III or Kugelberg-Welander disease, is mild with a variable clinical course and normal lifespan; commonly, this group is sub-divided into 3a with an onset between 18 months and 3 years old, and 3b with onset between 3 and 30 years old. Patients manifest muscular weakness with joint contractures and scoliosis but do not develop significant respiratory problems. They achieve independent walking; however, lose ambulation overtime. These individuals often have 3 to 4 copies of the *SMN2* gene. In SMA type IV, the affected have 4 or more copies of the *SMN2* gene; their symptoms begin after 30 years of age with a variable clinical course and mild muscular weakness; these patients have a normal lifespan and usually retain their independent walking ability. (Scheffer et al., 2001; Farrar et al., 2015; Calucho et al., 2018; Schorling et al., 2020; Keinath et al., 2021). Additionally, in recent years a rare new phenotype (less than 1% cases) has been identified. Type 0, with only one copy of SMN2, has a prenatal onset, showing reduced or absent foetal movement. At birth, affected individuals display hypotonia, contractures, severely deficient respiration, and motor functions, including lack of head control. These patients frequently die within weeks after birth (Keinath et al., 2021; Mercuri, 2021)

SMA type	Percentage of cases	Age of onset	Milestone	Life expectancy	SMN2 copies
0	<1%	Prenatal	Non-sitter, No head control	Peri-natal death	1
I	50-60%	0-6 months	Never roll or sit unaided	12 months to 24 months old	1-2
II	30%	6-18 months	Sitter, Never walk	Reach adulthood; 25-50 years	2-3
III	10%	After 18 months: IIIa: before 3 years IIIb: after 3 years	Walk unaided, loss of ambulance	Normal lifespan	3-4

IV	5%	Adulthood, >30 years	Walk unaided, progressive proximal weakness	Normal lifespan	4 or more

Table 1.1. Classification of Spinal Muscular Atrophy types.

1.4.1.1. SMA diagnosis

Generally, clinical diagnosis of motor neuron diseases begins with the assessment of clinical signs such as hypotonia and muscular weakness. Differential diagnosis first focuses on *SMN1* gene associated SMA, due to its frequency. This is usually confirmed with genetic testing, by detection of homozygous deletions of exons 7 and 8 in *SMN1* gene with a 95% of sensitivity. In case of hemizygous deletions or subtle mutations, quantitative assays of *SMN1* and *SMN2* exon7 can be performed to determine the number of gene-copies as well as mutation and linkage analysis (Scheffer et al., 2001; Mercuri et al., 2018). Further, diagnosis of SMA in new-borns is carried out on average around the age of 6 months, and choice of treatment depends on the copies of *SMN2* present in the patient. Some experts consider appropriate to initiate immediate treatment in children with one, two or three copies of SMN2, with or without symptoms, to prevent loss of motor neurons. Conversely, in case of no homozygous deletion/mutation of the *SMN1* gene in the affected, the evaluation continues with the assessment of creatinine kinase levels, and neurological studies, to help to discriminate among motor neuron diseases, neuromuscular junction disorders or myopathy (Schorling et al., 2020; Keinath et al., 2021; Nicolau et al., 2021).

As mentioned previously, there is a series of SMA disorders that are not associated to the *SMN1* gene, oftentimes called SMA 'plus' syndromes (Table 1.2). These variants display additional features to the motor neuron dysfunction, which may include, abnormalities of extraocular movement, arthrogryposis, deafness, encephalopathy, among others. Moreover, the SMA 'plus' syndromes can exhibit diverse patterns of muscular weakness which helps to differentiate them from *SMN1*-SMA. Proximal predominant disorders such as SMA with pontocerebellar hypoplasia, SMA with progressive myoclonic epilepsy, X-linked SMA, and SMA caused by mitochondrial

dysfunctions, are commonly the first considered during differential diagnosis when *SMN1*-SMA is discarded. Next, syndromes with distal muscular weakness predominance including SMA with lower extremity predominant (SMALED), Congenital distal SMA (DSMA), Scapuloperoneal SMA, SMA with respiratory distress (SMARD), and bulbar SMA syndromes (Brown Vialetto-Van Laere syndrome; Kennedy's disease) are contemplated. Finally, ALS and other neurodegenerative pediatric disorders such as Infantile neuroaxonal dystrophy (INAD), Achalasia-Addisonianism-Alacrima syndrome and Chèdiak-Higashi syndrome should also be considered (Farrar et al., 2015; Teoh et al., 2017).

Type of SMA 'plus' syndrome	Age of onset	Clinical features	Gene	Mode of inheritance
Pontocerebellar hypoplasia type 1 SMA	Early infancy	Diffuse weakness, severe hypotonia, central visual impairment, microcephaly, arthrogryposis	*EXOSC3* *EXOSC8* *SLC254A6* *VRK1*	Autosomal recessive
SMA with progressive myoclonic epilepsy	Childhood	Proximal muscular weakness, hypotonia, tongue fasciculation, probable hearing loss, facial weakness. Later myoclonic epilepsy	*ASAH1*	Autosomal recessive
X-linked SMA type 2	Prenatal	Similar to SMA type1, severe hypotonia, arthrogryposis	*UBE1*	X-linked
Cardioencephalo-myopathy with cytochrome C oxidase deficiency	Infantile	Similar to SMA type1, impaired extraocular movements, hypertrophic cardiomyopathy, seizures. Mitochondrial function and structure affected	*SCO2*	Autosomal recessive
Mitochondrial depletion syndrome 2	Infantile	Hypotonia, muscle weakness, respiratory failure, ophthalmic damage, seizures. Variable progression	*TK2*	Autosomal recessive
Mitochondrial depletion syndrome 3	Infantile	SMA with liver dysfunction, cerebral atrophy, and early death	*DGUOK*	Autosomal recessive

SMA lower extremity predominant (SMALED1)	Congenital to adult	Proximal predominant leg weakness, normal arm strength, nonprogressive	*DYNC1H1*	Autosomal dominant
SMALED2	Congenital to adult	Proximal and distal weakness of legs, mild weakness of arms with some contractures, slow progression	*BICD2*	Autosomal dominant
Congenital distal SMA (DSMA)	Congenital	Proximal and distal leg weakness, arthrogryposis. nonprogressive	*TRPV4*	Autosomal dominant
Scapuloperoneal SMA (SPSMA)	Early adult	Distal and scapuloperoneal weakness, laryngeal palsy, possible sensorineural deafness	*TRPV4*	Autosomal dominant
SMA with respiratory distress (SMARD)	Infancy	Diaphragm weakness, predominant distal extremities weakness	*IGHMBP2*	Autosomal recessive
SMA with respiratory distress 2 (SMARD2)	Neonatal	Distal muscular weakness, early diaphragm weakness, respiratory failure	*LASIL*	X-linked recessive
Brown-Vialetto-Van Laere (BVVL) syndrome	Childhood to adulthood	Pontobulbar palsy, ataxia, dysphagia; weakness of arms, hands, and face; fasciculations, sensorineural deafness. Progressive	*SLC52A3* *SLC52A2* *UBQLN1*	Autosomal recessive
Kennedy's disease	Adult	Progressive proximal and distal extremities and bulbar muscle weakness and atrophy; Prominent fasciculations, dysphagia, androgen resistance	*Androgen receptor*	X-linked recessive

Table 1.2. Spinal Muscular Atrophy 'plus' syndromes.

1.4.2. Genetic basis of SMA

SMA is an autosomal recessive disease caused by homozygous mutation or deletion of the SMN1 gene, located on the 5q13 chromosome region. (Scheffer et al., 2001) Approximately 95% of SMA cases are caused by loss of the gene, and 5% by mutations that produce a structural/functional deficient SMN isoform. Studies in different

populations have shown that Asians possess the most elevated carrier frequency of *SMN1* deletions with a 2.4% (Li, 2017; Beattie et al., 2018; Mercuri et al., 2018).

SMN2 (centromeric) is the second gene that produces the SMN protein. This is a gene highly homologous to *SMN1* (telomeric), originated by a duplication event in the 5q13 region and is located centromeric to the SMN1 locus. (Jangi et al., 2017). Nevertheless, a silent C/T transition in exon 7 of SMN2 disrupts a splice enhancer sequence, that results in omission of exon 7 in most SMN2-derived messenger ribonucleic acid (mRNA) transcripts, producing a truncated protein isoform SMNΔ7 that is rapidly degraded. As mentioned previously, the amount of full-length SMN protein produced by *SMN2* accounts for only about 10% of normal levels, which is enough for survival since complete loss of SMN protein is lethal; but results in SMA (Ahmad et al., 2016; Farrar et al., 2015; Lanfranco et al., 2017). Furthermore, the copy number variation of *SMN2* seem to inversely correlate with the clinical severity of the disease (Figure 1.14). Therefore, as a modifier of the disease *SMN2*, has been considered a target for therapeutic development to enhance the production of the full-length protein. (Beattie et al., 2018; Coady et al., 2011; Farrar et al., 2015; Jangi et al., 2017).

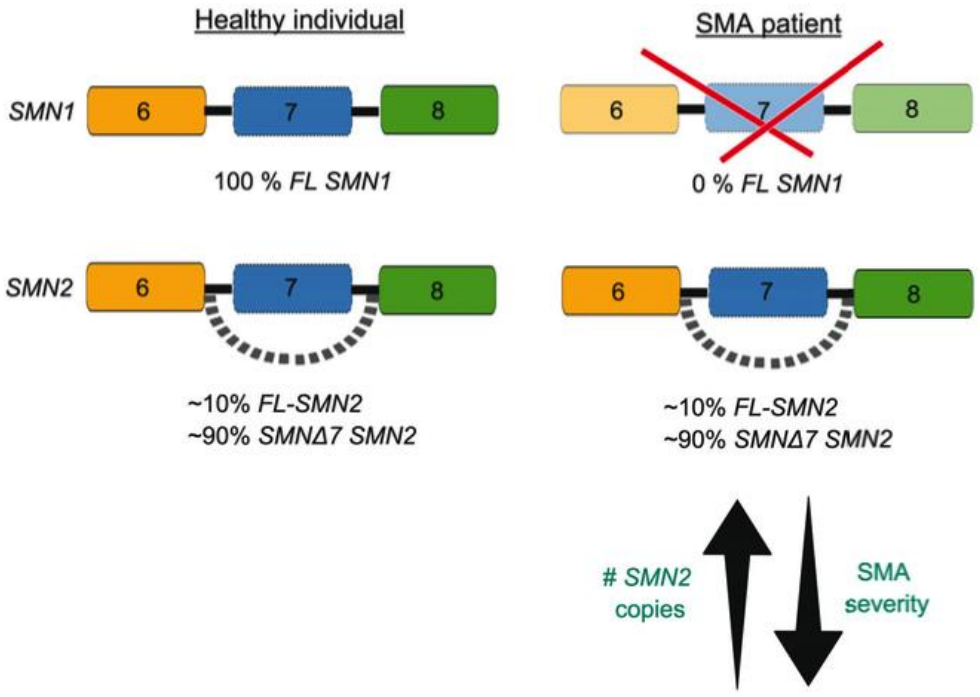

Figure 1.14. SMN1 and SMN2 in Spinal Muscular Atrophy In healthy individuals SMN1 and SMN2 genes produce 100% and about 10% of full-length SMN protein respectively. The 90% of the protein produced by SMN2, lack exon 7 due to aberrant alternative splicing, creating a non-functional protein. In SMA, the SMN1 gene is deleted or mutated; the remaining 10% of the full-length protein produced by SMN2 is enough for survival but translates into SMA, which severity correlates inversely with the copy number of SMN2 (From Bowerman et al., 2017).

1.4.3. SMN protein

SMN is a protein ubiquitously expressed, that has "housekeeping" and cell-type specific activities. It is highly present in brain, spinal cord, liver, and kidneys. This protein localizes in the cytoplasm and nucleus of all cells, accumulating in nuclear bodies called gems, frequently adjacent or overlapping Cajal bodies. (Coady et al., 2011; Jangi et al., 2017; Li 2017; Renvoisé et al., 2006)

SMN is encoded by 9 exons, has 294 amino acids and a molecular weight of 38kDa (Figure 1.15) The function of the SMN domains has been determined based on cellular assays and point mutations detected in SMA patients. On the Exon 2a 2b region, there is an N-terminal nucleic acid binding domain that overlaps the area of interaction with Gemin2, and mediates self-association (Coady et al., 2011). A Tudor domain is codified by Exon 3, this domain consists of 5 β-sheets that create an "aromatic cage" that recognizes and binds to symmetrically dimethylated arginine residues in numerous proteins including the Sm core proteins, and the Cajal Body marker coilin. (Coady et al., 2011; Raimer et al., 2017; Renvoisé et al., 2006; Zhao et al., 2016). Exon 5 contains a poly-proline motif that associates with profilin, an actin-binding protein. The Y/G box is a C-terminal domain that mediates oligomerization. Finally, Exon 7 provides stability and contains a QNQKE motif that targets proteins to the cytoplasm (Coady et al., 2011).

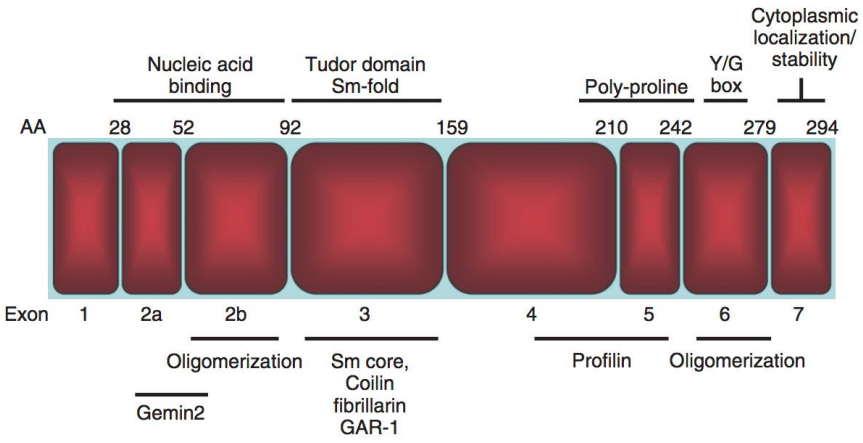

Figure 1.15. Survival motor neuron subdomains. SMN protein has 294 aminoacids. Relative size of exons and exon number are shown, as well as amino acid number corresponding to the exon peptides. The protein subdomains and associated functions are indicated, and below, the proteins that interact with SMN in the approximate site of interaction (From Coady et al., 2011).

SMN, is known to form a complex *in vivo* with Gemins 2-8 and Unr-interacting protein (Unrip) (Figure 1.16). Gemin2, Gemin4, Gemin8 and SMN can self-associate which creates large macromolecular SMN-Gemins complexes that cluster into nuclear bodies called gems (called U bodies in the cytoplasm). (Lanfranco et al., 2017)

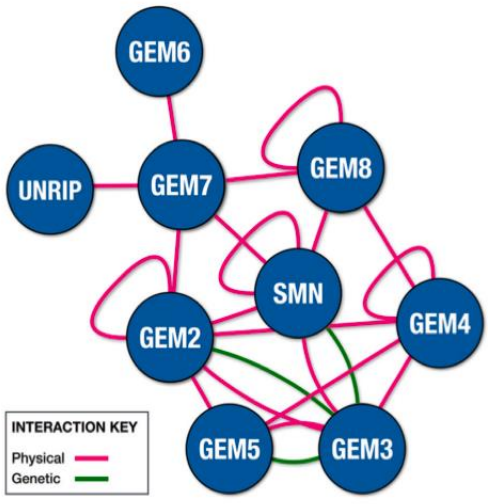

Figure 1.16. SMN complex. The SMN complex is formed by the association of SMN with the Gemins proteins 2-8. Physical and genetic interactions between the members of the complex are shown (From Lanfranco et al., 2017).

1.4.3.1. SMN protein functions

The survival motor neuron protein has a wide spectrum of functions from "housekeeping" activities through its canonical role in the assembly of Ribonucleoproteins as well as participation in metabolism of RNA and motor neuron development, among others (Figure 1.17) (Jangi et al., 2017; Beattie et al., 2018). The best understood role of SMA is its involvement in pre-mRNA splicing. Here, the SMN complex functions as a chaperone in the formation of small nuclear ribonucleoproteins (snRNPs), which are key elements of the splicing machinery. These ribonucleoprotein complexes consist in a ring of seven Sm (small RNA-binding) proteins (B/B', D1, D2, D3, E, F and G) around a single copy of a U-rich small nuclear RNA (U1, U2, U4, U5, U6, U11, U12, U4atac, and U6atac), and additional proteins that are specific for each snRNP. Once assembled, the snRNPs catalyze the removal of noncoding introns from pre-mRNAs during splicing (Beattie et al., 2018; Farrar et al., 2015; Jangi et al., 2017; Lanfranco et al., 2017).

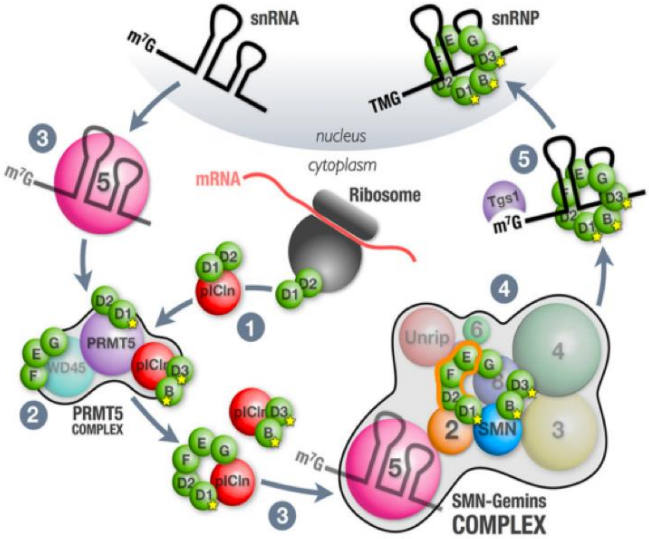

Figure 1.17. snRNP biogenesis. In the nucleus RNA pol II transcribes Sm-class snRNAs precursors that contain a m7G cap structure at the 5'-end and extra nucleotides at the 3'-end. **1)** Sm proteins are released from the ribosome, **2)** and associate with the PRMT5 complex. **3)** Pre-snRNAs exported to the cytoplasm, and Sm proteins are delivered to the SMN complex, **4)** Assembly of the Sm core. **5)** The snRNA is hypermethylated by Tgs1, and then is imported back to the nucleus (From Lanfranco et al., 2017).

Non-canonical functions proposed for SMN include chaperoning the transport of mRNA transcripts through motor neuron axons by associating with RNA-binding proteins from the transport machinery such as *Gap43* and *cpg15* in animal models (Donlin-Asp et al., 2016). SMN is also involved in actin-dynamics, more specifically the actin cytoskeleton at growth cones, through its interaction with profilin2a, a member of the Rho-kinase pathway (Figure 1.18) (Nölle et al., 2011; Hensel and Claus, 2018). Moreover, SMN is linked to ubiquitin homeostasis by impacting the Ubiquitin-like modifier 1 (Uba1) protein subcellular distribution and its spinal cord levels in mice (Wishart et al., 2014). Further, SMN is necessary for the maintenance of bioenergetic pathways and mitochondrial health in motor neurons and vesical release at neuromuscular junction (NMJ) (Xu et al., 2016; Bowerman et al., 2017). More recently, it has been suggested that SMN protein is also involved in R-loop resolution during RNA pol II transcription termination. The SMN Tudor domain recognises and binds to the symmetrically dimethylated arginine residue (R1810) from the carboxy-terminal domain (CTD) of RNA pol II; and interacts with SETX a helicase involved in R-loop resolution during transcription termination (Jangi et al., 2017; Zhao et al., 2016).

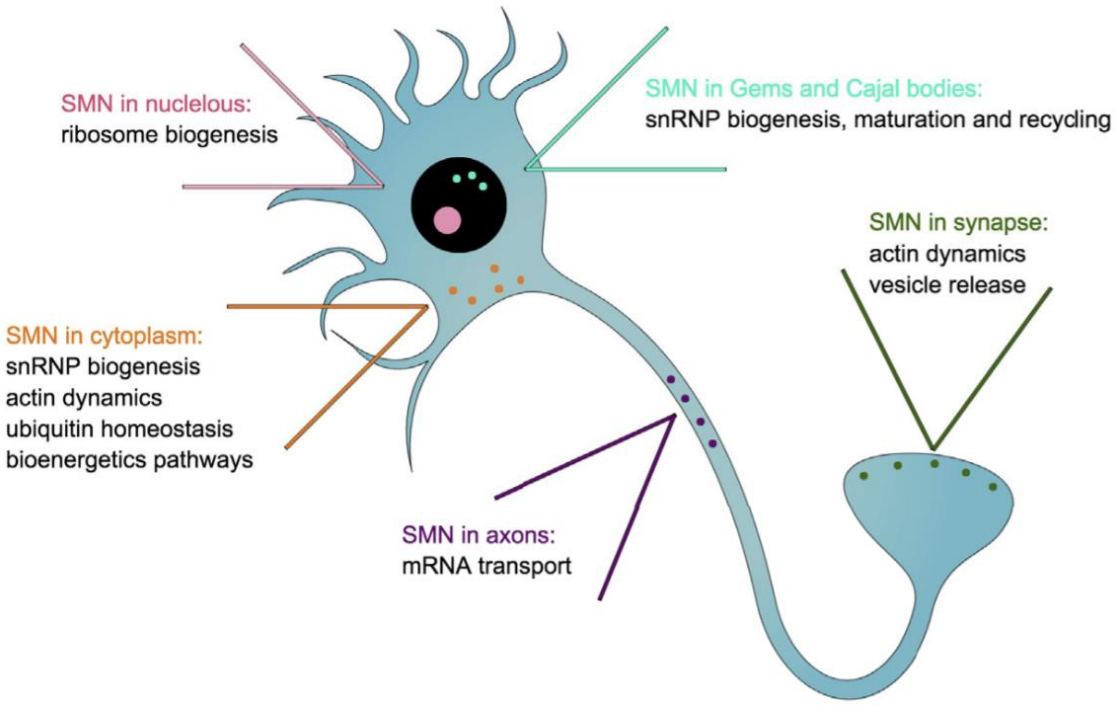

Figure 1.18. SMN functions. Schematical representation of some of the functions that have been proposed for SMN in neurons (From Bowerman et al., 2017).

1.4.4. SMA pathophysiology

Loss of SMN creates a variety of perturbations over the human body, including the disturbance in the levels/functions of proteins caused by splicing defects or by lack of interaction with SMN. The most representative is the specific vulnerability of motor neurons. Several mechanisms for neurodegeneration have been proposed; however, there is much left to uncover since disruption of the canonical function of SMN in splicing generates more 'systemic' defects. Meaning that, deficient chaperoning of the SMN complex during the snRNPs assembly, causes pre-mRNA splicing defects through a decrease in snRNPs. (Zhang et al., 2008; Bowerman et al., 2018). Further SMN is necessary to prevent widespread intron retention during splicing events. This hinders transcription and creates DNA damage and cellular stress due to R-loops formed over retained introns (Jangi et al., 2017).

Among the suggested mechanisms that participate in the selective degeneration of neuronal cells in SMA, their bioenergetic profile has been included due to their high energetic demand. Induced pluripotent stem cells from SMA patients that were later differentiated into motor neurons displayed a diminished area and number of mitochondria as well as deficient mitochondrial transport in axons (Xu et al., 2016). Further, knockdown of phosphoglycerate kinase 1 (Pgk1) resembles the SMA phenotype on zebrafish and increasing biogenesis of the mitochondria rescued abnormal motor axons in SMA zebrafish (Boyd et al., 2017). Motor neuron development also appears to be affected in SMA. SMN interacts with the member of the Rho-kinase (ROCK) pathway, profilin2a, which is hyper-phosphorylated in SMA cells. Neurite outgrowth in PC-12 cells is dependent on profilin2a phosphorylation and inhibition of the ROCK pathway rescued the neurite growth defect by SMN loss (Nölle et al., 2011).

Moreover, a variety of mRNAs have been found to be mislocalised in SMA motor neurons, together with reduced levels of SMN-interacting proteins necessary for the localization and translation of mRNAs. The zipcode binding protein (ZBP1 or IMP1) interacts with the RNA-binding protein HuD from the ELAV family, to modulate the localization of *cpg15* and growth associated protein 43 (*Gap43*) transcripts, which are mislocalised in SMA motor neurons (Donlin-Asp et al., 2016). Other processes like

synapsis and ubiquitination defects have also been found in SMA. Neuromuscular junctions in SMA mice present aberrant synaptic transmission with lower vesicle density and reduced vesicle release (Kong et al., 2009). Uba1 expression in SMA mice is reduced, which results in disruption of ubiquitin homeostasis, myelination defects in Schwann cells and motor neuron degeneration (Aghamaleky Sarvestany et al., 2014).

Further, a number of factors associated with DNA damage are differentially regulated in SMN-deficient cells. Experiments carried out in a SMNΔ7 mouse model revealed lower expression of the Smn protein in medial motor column motor neurons labelled as vulnerable; and these cells showed accumulation and activation of p53, a transcription factor that gets activated during cellular stress and DNA damage. This p53 phosphorylation promotes selective death of the vulnerable motor neurons (Simon et al., 2017). Moreover, upregulation of the p53 downstream factor and mediator of cell cycle arrest, cyclin-dependent kinase inhibitor Cdkn1a/p21 was also observed in motor neurons and spinal cord of animal models. However, recent studies on a milder mouse model of SMA ($Smn^{2B/-}$) showed that depletion of p53 and p21 do not impede motor neuron loss; thus, they are not the main drivers of motor neuron death (Jangi et al., 2017; Reedich et al., 2021).

SMA has also been associated with R-loop mediated DNA damage; chronic low levels of SMN cause the downregulation of the R-loop helicase SETX and DNA-dependent protein kinases (DNA-PKcs). This generates accumulation of R-loops and partial impairment of the NHEJ DNA repair pathway (Kannan et al., 2018; Cuartas and Gangwani, 2019). Further, the zinc finger protein ZPR1, a factor that upregulates SMN expression and is necessary to its location to nuclear gems and cajal bodies, is downregulated in SMA patients. ZRP1 binds to RNA Pol II and form complexes with SETX to remove R-loops. The overexpression of ZPR1 in SMA mice reduces neurodegeneration; hence, presenting itself as a potential modifier for SMA therapy (Kannan et al., 2020; Kannan et al., 2022).

1.4.5. Management of SMA patients.

The standards of care for spinal muscular atrophy include diagnosis, physical therapy, nutrition, medication, palliative care among others (Finkel et al., 2018; Mercuri et al., 2018). The overall management of these patients require a multidisciplinary approach in collaboration with the family. Regular physical examinations provide information about the muscular strength status, which is the basis for the design of physical therapy sessions. Orthopaedic management include assessments to determine if patients are candidates for surgical intervention for spinal deformities, hip instability, or fractures. Moreover, organs and systems are also monitored in patients, since they can suffer from swallowing and gastrointestinal dysfunction or pulmonary ventilation problems; as well as bone health in the presence of osteopenia (Finkel et al., 2018; Mercuri et al., 2018).

Several drug treatments for SMA have employed molecules like creatine, gabapentin, valproic acid, thyrotropin-releasing hormone, albuterol, among others. Unfortunately, these therapies have not shown significant improvement on the disease course so far. Moreover, complementary treatment for symptoms comprises antibiotics, supplements, and annual immunizations against influenza and *pneumococcus*, which are strongly recommended (Finkel et al., 2018; Schorling et al., 2020).

Currently therapy development is centred on splicing modifiers for SMN2 gene, replacement of SMN1 gene or enhancing muscle growth/function (Figure 1.19). The antisense oligonucleotide (ASO) drug called nusinersen, has been approved as a treatment for all SMA types. This drug facilitates the inclusion of SMN2 exon7, by binding to the splice-silencing-site in intron 7; thus, generating a higher number of SMN2-mRNA with exon7 incorporated and ultimately producing a functional full-length protein (Schorling et al., 2020; Nicolau et al., 2021). Other splicing modifiers that are being investigated in trials include the small molecules risdiplam (RG7916) and branaplam (LMI070) which also aim to increase the levels of full-SMN protein. The gene therapy approach delivers a wild-type copy of SMN through an Adeno-associated viral serotype 9 (AAV9) vector intravenously. The FDA approved the use of Zolgensma (AVXS-

101) on patients below 2 years of age in 2019 (Schorling et al., 2020). Finally, neuroprotective drugs like Olesoxime, an apoptosis inhibitor is still under study; along with Fast skeletal Muscle troponin activators (FSTA) and Myostatin-inhibitors that promote muscle mass growth and function (Finkel et al., 2018; Schorling et al., 2020).

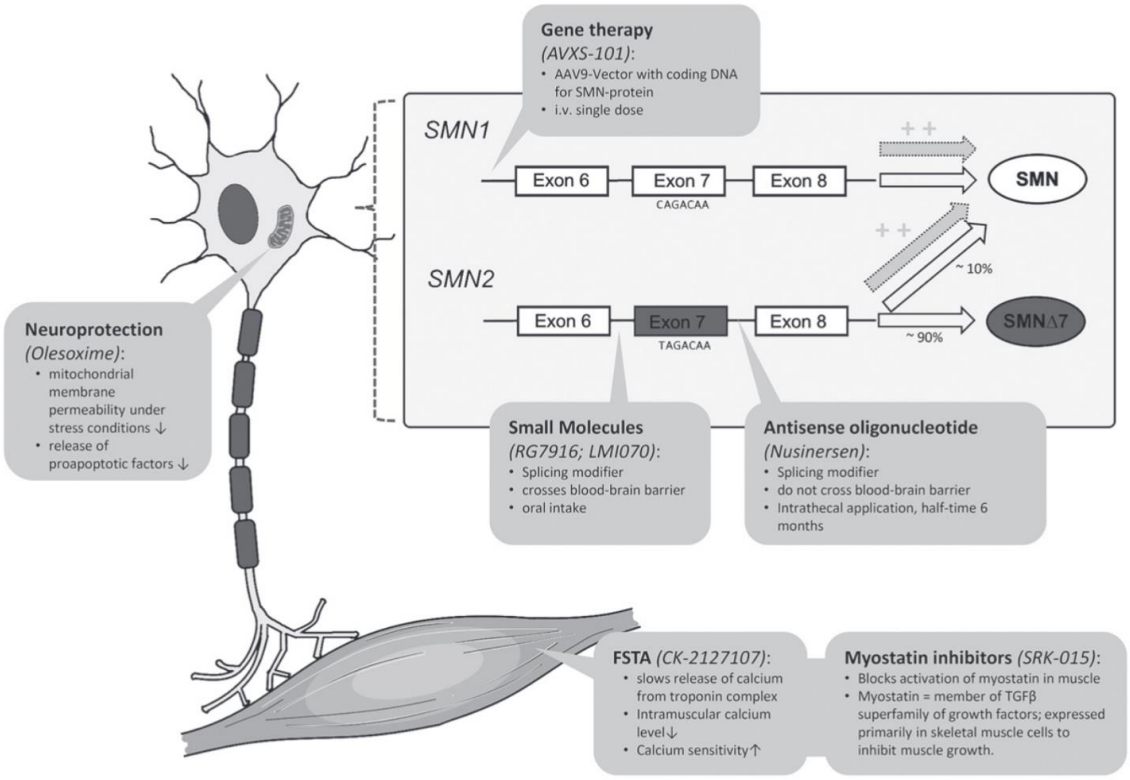

Figure 1.19. SMA therapeutic approaches. Different therapies developed for the treatment of Spinal muscular atrophy (SMA). Nusinersen and the AAV9 vector (AVXS-101) have been approved for human use. The remaining approaches are still being studied. FSTA: Fast troponin activator (From Schorling et al., 2020).

2. CHAPTER II. MATERIALS AND METHODS

2.1. SOLUTIONS

Acetonitrile (ACN) 50%

50 mL of Mass Spectrometry (MS) grade ACN plus 50 mL of MS grade were mixed and kept at room temperature (RT).

Acetonitrile (ACN) 70%

70 mL of MS grade ACN plus 30 mL of MS grade were mixed and kept RT.

Acrylamide 30%

Ultra-pure ProtoGel® 30% (w/v) Acrylamide: 0.8% (w/v) Bis-Acrylamide stock solution (37.5:1).

Ammonium Bicarbonate (AB) 100 mM

0.395 g of AB (MW: 79.06), dissolved in 50mL of Mass Spectrometry grade H_2O. Made fresh each time. Kept at RT.

Ammonium Persulphate (APS) 10%

APS was dissolved in dd H_2O w/v. 500µL aliquots kept at -20°C.

XX REDACTED TEXT XX

XX REDACTED TEXT XX

BSA 3% in PBS

0.6 g of Bovine Serum Albumin (BSA) were added to 20 mL PBS buffer pH7.4. Solution was filter-sterilized with a 0.22 µm filter. Made fresh each time.

BSA 3% in TBST

0.6 g BSA were dissolved in 20 mL TBST buffer.

Calcium Chloride $CaCl_2$ 2.5M

27.7 g $CaCl_2$ $2H_2O$ - MW=110.99, complete to 100 ml with ddH_2O.

ChIP lysis Buffer

50 mM HEPES-KOH pH 7.5, 140 mM NaCl, 1 mM EDTA pH 8, 1% Triton X-100, 0.1% Sodium Deoxycholate, 0.1% SDS; 1X Protease inhibitors (fresh each time), Roche.

DRIP Binding buffer 10X

100mM sodium phosphate pH7 was mixed with 1.4M NaCl and 5% (v/v) Triton X-100. The solution was filter sterilized with a 0.22 µm filter. To prepare 1X DRIP binding buffer, the previous solution was diluted ten times in TE buffer, Invitrogen™. Kept at room temperature up to a month.

DRIP elution buffer

50mM Tris pH8 was mixed with 10mM EDTA pH8, and 0.5% (v/v) SDS. Kept at RT.

Elution Buffer

1% SDS, 100 mM $NaHCO_3$. Kept at room temperature (RT).

Ethylenediaminetetraacetic acid (EDTA) 4%

27.397mL of 0.5M EDTA, top up to 100 mL with sterile ddH_2O w/v. Kept at 4°C.

EDTA- Hanks' balanced salt solution (HBSS)

50 μL of EDTA 0.5 M were added to 50 mL of HBSS, solution was filter sterilized using a 0.22 μm filter, kept at RT.

Formic Acid (FA) 0.5%

500 μL FA plus Q.S. (quantum satis) 100 mL MS grade H_2O, kept at RT.

Formic Acid (FA) 10%

10 mL FA plus 90 mL MS grade H_2O, kept at RT.

Glycerol 80%

>99% Glycerol, SIGMA® diluted with dd H_2O, kept at RT.

Guanidine HCl (GndHCl) 6M

573 mg GndHCl (MW: 95.53 g/mol) plus Q.S. 1mL 100 mM AB, kept at RT.

HBSS Buffer 2X

280 mM NaCl, 100mM HEPES, 1.5mM Na_2HPO_4; Adjust the pH to 7.1 with NaOH, topped up to 500ml with ddH_2O. Stored at –20°C in 50 mL aliquots.

High Salt Wash Buffer

Add 0.1% SDS, 1% triton X-100, 2 mM EDTA, 20 mM Tris-HCl pH 8.0, 300 mM NaCl in ddH_2O, kept at RT.

Iodoacetamide (IAA) 50mM

9.25 mg IAA (MW: 184.96) plus 1 mL of 100 mM AB. Solution was prepared fresh, in the dark and covered with aluminium foil, after use the rest was discarded.

LB Agar

14 g of LB Agar powder dissolved in 400 mL of ddH2O. Autoclaved 121°C for 15 minutes, kept at RT.

LB Broth

14 g of LB Broth powder dissolved in 400 mL of ddH2O. Autoclaved, kept at RT.

Lithium Chloride Wash Buffer

Add 0.25 M LiCl, 1% NP-40, 1% Sodium Deoxycholate, 1 mM EDTA, 10 mM Tris-HCl pH 8.0 in ddH_2O, kept at RT

Low Salt Wash Buffer

0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl pH 8.0, 150 mM NaCl in in ddH_2O, kept at RT.

Lysis base buffer

20 mM HEPES pH 7.4, 40 mM NaCl 2mM $MgCl_2$, and 1% Triton X-100 in ddH_2O. Kept at 4°C.

Methanol: Acetone solution 1:1

25 mL of 100% methanol, mixed with 25 mL of 100% acetone in the fume hood. 50mL tube was kept at -20°C.

Paraformaldehyde (PFA) 4%

16% PFA solution diluted in PBS, kept at 4°C. Alternatively, a 10% formalin solution (equivalent to 4% PFA) was used, and kept at RT.

Paraformaldehyde (PFA) 37%

1.85g PFA, 3.5mL ddH_2O, 10µL 10M KOH heated in the microwave in 3 second lapses until dissolved. Kept in fume hood, used immediately and the rest was discarded.

PBS buffer 1X (Phosphate buffered Saline)

1 PBS tablet, in 500mL of ddH_2O. Autoclaved, kept at RT.

PBS/tween 20 0.02%

Tween 20 diluted in PBS buffer v/v. Kept at RT.

Poly-D-lysine (5X stock)

Lyophilized poly-D-lysine was hydrated in the hood with 50 mL of dH2O and mixed well. The bottle was kept at -20°C. Further dilution with dH2O was needed to prepare of 1X solution.

Protein Loading Buffer (PLB) 5X

1g SDS powder, 771.25mg DTT powder, 2.5ml 250mM Tris-HCl pH6.8 solution, 500mg Bromophenol Blue powder, 5ml 50% (w/v) Glycerol and 2.5ml ddH2O, mixed in the fume hood. Kept at RT.

RIPA buffer

Add 50 mM Tris-HCl pH 8, 150 mM NaCl, 2 mM EDTA pH8, 1% NP-40, 0.5% Sodium Deoxycholate, 0.1% SDS, 1X Protease inhibitors (fresh each time) in ddH_2O. Kept at RT protected from light.

Running buffer 10X

187.7 g Glycine, 30.3 g Tris Base, 10 g SDS, ddH_2O up to 1L. Kept at RT.

Sodium deoxycholate 10%

2g Sodium deoxycholate, 20 mL ddH_2O, incubated at 37°C vortexing every 5 minutes until dissolved. Kept at RT protected from light.

Sodium Dodecyl Sulphate (SDS) 2% and 10%

Using a face mask, the corresponding amount of SDS powder was weighed and dissolved in ddH_2O w/v. Kept at RT.

Sodium Phosphate 1M pH7

39 mL of 2M sodium phosphate monobasic, SIGMA®; were mixed with 62 mL of 2M sodium phosphate dibasic heptahydrate, SIGMA®; and 100mL of RNase-free water. Kept at RT.

Tris base 50mM

Tris powder dissolved in ddH_2O adjust pH to 7.4 with HCl. Kept at RT.

TBE buffer 10X (Tris Borate EDTA buffer)

108 g Tris base, 55 g Boric Acid, 40 mL of 0.5 M EDTA, ddH_2O up to 1L. Kept at RT.

TBS buffer 10X (Tris Buffer Saline)

24.2 g Tris base, 80 g Sodium Chloride, ddH_2O up to 1L, pH 7.9. Kept at RT.

TBST 1X (Tris Buffer Saline Tween)

100mL TBS, 1 mL Tween 20 SIGMA®, ddH_2O up to 1L. Kept at RT.

TE buffer

Thermo Fisher Scientific Ph 8.0, kept at RT.

Transfer Buffer 1X (semi-dry transfer)

200mL of 5X Trans-Blot® Turbo™ transfer buffer and 200mL of 100% ethanol; diluted with 600mL of ddH_2O. Kept at room temperature.

Tris (2-carboxyethyl) phosphine (TCEP) 50 mM

100 μL 0.5M TCEP SIGMA® dissolved in 900 μL of 100mM AB. Ready-to-use aliquots were kept at -20°C.

Trypsin 0.1 μg/μL (MS)

20 μg of trypsin powder dissolved in 200 μL of 100 mM AB. Ready-to-use aliquots were kept at -20°C.

Trypsin 0.25%

Trypsin powder dissolved in ddH_2O, w/v, filter sterilized, kept at 4°C.

Trypsin-EDTA solution

0.1% Trypsin, 0.4% EDTA in PBS buffer, kept at 4°C.

Trypsin-EDTA solution 1X

10X trypsin, SIGMA-ALDRICH®; diluted in PBS buffer. Kept at 4°C.

Wet transfer Buffer 1X

100 mL of 10X Tris-Glycine Transfer buffer, G-biosciences; 200 mL of methanol and 700 mL of ddH_2O. Kept at RT.

2.2. TISSUE CULTURE

Below I present a brief description of the cell lines used in this research (Table 2.1) and the methodology for their maintenance. Some experiments were carried out in immortalised cell lines which are usually easy to work with and give good results. However, in a few cases I chose to work with primary fibroblasts from SMA patients and healthy individuals to recreate as much as possible the context of the disease, since the depletion of SMN expression by synthetic agents such as shRNAs or microRNAs was not efficient enough. Of note, SMN is produced by two genes *SMN1* and *SMN2* and is possible that this hindered the regulation of protein expression by the synthetic molecules.

2.2.1. Media preparation

2.2.1.1. Maintenance media

Media bottles and supplements were first warmed up to 37°C in the water bath, next they were sprayed with 70% IMS, wiped with tissue paper, and placed inside the laminar flow hood. 50mL of media was discarded from the original bottle and the same volume of foetal calf serum (FCS) was added to the bottle, together with the rest of the supplements. Media was mixed and kept at 4°C.

2.2.1.2. XX REDACTED TEXT XX

XX REDACTED TEXT XX

2.2.1.3. iPSc derived Motor Neuron media (iPSc derived MN differentiation and expansion performed by Dr. Vinay Kumar Godena, SITraN).

All medias were prepared to final volume of 50 mL. The basic media was prepared using DMEM/F12 – Neurobasal (1:1; 25 mL/25 mL) Gibco™; supplemented with N2 (0.5 X; 250 μL); B27 (0.5 X; 250μL); Ascorbic acid (0.1 mM; 22 μL of 40 mg/mL stock); Glutamax (1X; 500 μL of 100 X stock); and P/S (1X; 500 μL of 100 X stock).

For the differentiation process, on day 1 to 6 of the basic media was further supplemented with CHIR (3 μM; 7.1 μL of 21.08 mM stock); DMH1 (2 μM; 3.8 μL of 26.28 mM stock) and SB (2μM; 7.68 μL of 13 mM stock). On day 7 to 12, the additional supplements for the basic differentiation media were CHIR (1μM; 2.37 μL of 21.08 mM stock); DMH1 (2 μM; 3.8 μL of 26.28 mM stock); SB (2μM; 7.68 μL of 13 mM stock); RA (0.1 μM; 5 μL of 1 mM stock); and PMN (Pur) (0.5 μM; 25 μL of 1 mM stock).

For the expansion of MN, on day 13 to 18, the basic media was further supplemented with RA (0.5 μM; 25 μL of 1 mM stock); and PMN (Pur) (0.1 μM; 5 μL of 1 mM stock). On day 19 to 28 the additional supplements for the basic differentiation media were RA (0.5 μM; 25 μL of 1 mM stock); PMN (Pur) (0.1 μM; 5 μL of 1 mM stock); Compound-E (0.1 μM; 4.9 μL of 1.02 mM stock); BDNF (10ng/mL; 5μL of 100μg/mL stock); CNTF (10ng/mL; 5μL of 100μg/mL stock); and IGF-1 (10ng/mL; 5μL of 100μg/mL stock).

2.2.2. Cell subculture

2.2.2.1. HEK 293T

Human Embryonic Kidney cells 293 that express a mutant version of the SV40 large T antigen (HEK293T) cells were maintained in Dulbecco's Modified Eagle's Medium (DMEM), SIGMA-ALDRICH®; supplemented with 10%FCS, SIGMA-ALDRICH®; 1% pen/strep, Thermo Fisher; and 1% L-Glutamine, SIGMA®. Alternatively, for viral production, HEK293T cells were grown in DMEM, LONZA®; 10%FCS, SIGMA-ALDRICH®; and 1% pen/strep Thermo Fisher.

To split the cells, media was removed from a ~80% confluent flask. Next, cells were rinsed once with warm PBS, and a suitable volume of Trypsin-EDTA solution was added to the cells to dislodge them from the flask surface (2mL to T75 flasks and 4mL to T175 flasks). Cells were incubated for 5 minutes at 37°C and 5% CO_2; after, media was added to the cells to inactivate the trypsin (5mL to T75 flasks and 10 to T175 flasks). Cell suspension was collected into a universal tube and centrifuged at 180 xg for 5 minutes; pellets were resuspended in warm media and a specific volume of the suspension was transferred to the flask according to the subculture dilution needed (usually 1:8 or 1:10). An appropriate volume of media was added to the cells depending on the container size. (13 mL to T75 flasks and 25 mL to T175 flasks). Cells were kept at 37°C and 5%CO_2.

2.2.2.2. HeLa

Immortal cell line derived from Henrietta Lacks' cervical cancer tumour, was used for viral titration. For these experiments Eagle's Minimum Essential Medium (EMEM) SIGMA-ALDRICH®; 2mM Glutamine, SIGMA-ALDRICH®; 1% NEAA, SIGMA-ALDRICH®; 10% FBS, Gibco) was used as growth medium. Cells were split as described on 2.2.2.1 for HEK293T cells.

2.2.2.3. Human fibroblasts

Spinal muscular atrophy type 1 (SMA1) human fibroblasts (GM09677) and healthy human fibroblasts (GM08680) obtained from Coriell Institute (Table 2.1) were

maintained in Dulbecco's Modified Eagle's Medium LONZA® (DMEM) supplemented with 10% FCS, BIOSERA®; 1% Pen/strep, Thermo Fisher; and 1:1000 50µg/mL Uridine.

To split these cells, after removing the media from a T175 flask and washing once with warm PBS, 5mL of EDTA-HBSS solution was added to the cells and incubated for 2 minutes at 37°C and 5% CO_2. Next, the EDTA-HBSS solution was discarded and 5mL of Trypsin-EDTA 1X was added and the cells were incubated for 5 minutes. Trypsin was inactivated by adding 7 mL of full media. The cell suspension was then transferred to a new flask in a 1:3 ratio and topped with 25mL of fresh media.

2.2.2.4. Human Lymphoblasts

Human β-lymphocytes transformed with Epstein-Barr virus were used for viral titration. Cells were maintained in suspension using Roswell Park Memorial Institute Medium 1640 supplemented with 2mM L-glutamine, SIGMA®; and 10% Foetal Calf serum (FCS), BIOSERA®. These cells were split by dilution in a 1:4 ratio. First cells were resuspended by pipetting, next the desired volume of cell suspension was transferred to a new T75 flask standing upright and topped with fresh media to a 50 mL final volume.

2.2.2.5. iPSc - derived Motor Neurons

Induced pluripotent stem cells (iPSc)- derived Motor Neurons from SMA patients and healthy individuals, were differentiated and maintained according to a published method (Du et al., 2015). This resulted in a relatively pure population (>95%) of Motor Neuron Progenitors (MNPs), and enriched batches of functional Motor Neurons (MN) (>90%) that could be maintained for a minimum of 5 passages. MN were split 1:6 once a week, and frozen in regular media (DMEM/F-12 + FBS + 10% DMSO).

2.2.2.6. MRC5

Human lung fibroblasts cells from foetus, were maintained in Dulbecco's Modified Eagle's Medium (DMEM), SIGMA-ALDRICH®; supplemented with 10%FCS, SIGMA-

ALDRICH®; 1% pen/strep, Thermo Fisher; and 1% L-Glutamine, SIGMA®. The cells were split as described on 2.2.2.1 for HEK293T cells.

Cell line	Remarks
HEK 293T	Express a mutant version of the SV40 large T antigen. Female foetus. Hypotriploid
MRC5	14-week aborted human male foetus. Diploid.
Human fibroblasts Healthy	Coriell Institute GM08680 Male, 5 months old at sampling. Source: skin, foreskin
Human fibroblasts SMA type I	Coriell Institute GM09677 Male, expired at 23 months age. Source: lens, eye. Homozygous for Ex7-8del of SMN1 gene. Has 3 copies of SMN2 gene.
Human lymphoblasts healthy	Coriell Institute AG20934 Male, new-born. Source: blood, umbilical cord. Cells grow in suspension
Human lymphoblasts SMA type I	Coriell Institute GM23689 Male, 1 year at sampling. Source: blood, peripheral vein. Clinically affected.
iPSc derived motor neurons Healthy	Internal codes: CS14i; MIFF1 Isolated from healthy individuals.
iPSc derived motor neurons SMA	Internal codes: SMA86; SMA32 Isolated from SMA type I patients
HeLa	Hyper triploid chromosome number (3n+)

Table 2.1 Cell lines remarks. Summary of the cell lines used in this study.

2.2.3. Cell count and seeding

Cells were counted by placing 10μL of cell suspension onto a Neubauer chamber. The number of cells on the 4 corner quadrants was averaged and multiplied by the dilution factor $1x10^4$ (cells/mL).

After counting, the volume of cell suspension was calculated for seeding, warm media was added to the flask or plate and then the cells. Cells were incubated at 37°C and 5% CO_2. Table 2.2 shows the number of HEK293T cells seeded in different containers.

	Number of cells	Medium volume
24 well plate	30 x 10^3 cells/mL	1mL
12 well plate	75 x 10^3 cells/mL	2mL
6 well plate	2-3 x 10^5 cells/mL	3mL
10 cm dish	2-3 x 10^6 cells/mL	12mL
15cm dish	6-8 x 10^6 cells/mL	20mL

Table 2.2 Cell seeding density. Number of cells seeded in different size containers and volume of media used. Based on HEK293T cells.

2.2.4. Transfection

2.2.4.1. Plasmid Transfection

Cells were seeded on day 1. Transfection was carried out the next day, when cells reached 60% confluency, using polyethylenimine (PEI) as a transfection reagent in a 2:1 PEI:DNA ratio. On day 2, the transfection mix was prepared as follows for a 15 cm plate:

Tube A: 10 μg DNA in 300 μL of Serum-free media (SFM)

Tube B: 20 μL of PEI in 300 μL of SFM

First, in the hood, warm SFM was placed into 2 Eppendorf tubes; next, DNA and PEI were added to tube A and B correspondently, mixing gently by pipetting. After 5-minute incubation at RT, tube A was mixed with tube B by pipetting and incubated for 20 minutes at RT. Meanwhile, old media in the plates was replaced by new warm media;

later, 600 μL of transfection mix were added to the cells dropwise over the whole plate surface. Cells were incubated 48h at 37°C and 5% CO_2. For a 6 well plate 1.5 or 2 μg of DNA was used, and 100 μL of SFM per tube for a final volume of 200 μL of transfection mix.

2.2.4.2. siRNA Transfection

Cells were seeded to reach 60-70% confluency at transfection. Metafectene® or Lipofectamine® RNAiMAX were used as transfection reagents. The transfection mix was prepared as follows for a 6-well plate:

Metafectene® 1:1 v/v ratio, 50nM.

Tube A: 3 μL of 20 μM siRNA in 100 μL of Opti-MEM®

Tube B: 3 μL of Metafectene® in 100 μL of Opti-MEM®

Lipofectamine® RNAiMAX 1:3 v/v ratio, 15nM.

Tube A: 3 μL of 10 μM siRNA in 150 μL of Opti-MEM®

Tube B: 9 μL of Lipofectamine® RNAiMAX in 150 μL of Opti-MEM®

After preparing the transfection mix using the same incubation times and mixing techniques as the previous numeral stated, the old media in the wells was replaced by new warm media (1mL of media for Metafectene® transfections and 1.7 mL for the Lipofectamine® transfections). Finally, the transfection mix was added to the cells dropwise over the whole well surface, cells were incubated 48h at 37°C and 5% CO_2. In the case of transfections with Metafectene® media was topped up to 2mL 24h after transfection to decrease cell toxicity.

For a 15 cm plate using Metafectene®, reagents were diluted in 800 μL of Opti-MEM® and added to 8 mL of new media on the plates for a final volume of 9.6 mL, on the next day media was topped up to 20 mL to diminish cell toxicity.

2.2.4.3. XX REDACTED TEXT XX

XX REDACTED TEXT XX

2.2.4.4. XX REDACTED TEXT XX

XX REDACTED TEXT XX

2.2.5. Transduction

Viral transduction of fibroblasts was carried out in 6 well plates. First, fibroblasts in cell suspension were counted, and the correspondent volume containing 3 x 10^5 was transferred to a 15 mL falcon tube. Next, the volume of virus needed was calculated using the following formula:

$$Vol = \frac{MOI \times No.\ of\ cells}{Titre}$$

MOI: multiplicity of infection (optimized for each virus- in this case was 30 for both).

Titre: LV-GFP 4.7 x 10^8

LV-HB-GFP 9.5 x 10^7

The volume of virus calculated was added to the tube and the suspension volume was topped up to 3 mL with warm media. Suspension was then poured into a well of a 6 well plate. Cells were incubated 48 h at 37°C and 5% CO_2.

2.2.6. CPT treatment

Media was removed from cells growing in 15 cm plates, leaving only 10 mL. Next, Camptothecin (CPT) was added to a final concentration of 25 µM to the media and gently swirled; cells were incubated for 10 minutes at 37°C and 5% CO_2. After, media was removed, and cells were washed twice with cold PBS.

2.3. MOLECULAR BIOLOGY TECHNIQUES

2.3.1. Bacterial Transformation

Competent DH5α cells NEB®, were thawed on ice for 10-15 minutes until all ice crystals disappeared. Next, 50 µL of competent cells were placed into a transformation tube on ice, and 1-5 µL of plasmid DNA (1 pg-100 ng) were added to the cells and incubated on ice for 30 minutes. After incubation cells were heat shocked at 42°C for 30 seconds and incubated on ice for 5 minutes. Later, 950 µL of SOC outgrowth medium, NEB® were added to the cells and incubated at 37°C for 60 minutes at 250 rpm. Next, 200 µL of cell suspension were spread on a warm selection plate. Cell suspension was then centrifuged at 2500 rpm for 2 minutes; 600 µL of supernatant were discarded, and pellets resuspended. The remaining 200 µL of cell suspension were spread on a warm selection plate. Plates were incubated overnight at 37°C.

2.3.2. Clonal expansion

For mini preps, a colony from a selection plate was taken carefully, using a 200 µL micropipette tip and placed into a universal tube containing 5 mL of LB broth with the suitable antibiotic. Tubes were incubated overnight at 37°C/250 rpm.

For midi preps, single colonies were expanded using 50 mL of LB broth with the suitable antibiotic in conical flasks and incubated overnight at 37°C/250 rpm.

2.3.3. Preparation of Glycerol Stocks

750 µL of cell suspension were placed into a cryo-tube with a screw cap, 250 µL of 80% Glycerol were added and gently mixed. Kept at -80 °C.

2.3.4. Nucleic Acid Purification

QIAGEN® QIAprep Spin Miniprep Kit: For purification of up to 20 µg of Plasmid DNA.

QIAGEN® Plasmid Plus Midi Kit: For purification of up to 10 mg of Plasmid DNA.

QIAGEN® QIAquick PCR purification Kit: For purification of up to 10 µg of PCR products.

2.3.4.1. Phenol-Chloroform DNA extraction

One equal volume of UltraPure™ Phenol: Chloroform: Isoamyl Alcohol, ThermoFisher; was added to DNA solution. After mixing gently and spinning for 5 minutes at maximum speed, the aqueous layer (top layer) was removed and placed into a new tube. One equal volume of Chloroform SIGMA®, was added to the aqueous layer; the tube was mixed and spun down for 2 minutes at maximum speed. The aqueous layer was again removed and placed into a new tube.

For the DNA precipitation with ethanol, two volumes of 100% Ethanol were added to the top layer, plus sodium acetate NaOAc pH 5.2 to a final concentration of 0.3M, and 1 µL of Glycogen, Invitrogen™. Samples were mixed and frozen at -80°C for 30-45 minutes then spun down for 30 minutes at maximum speed and 4°C; alternatively, samples were left overnight at -20°C, and then spun as previously described. The supernatant was removed, and the pellet was washed with 70% Ethanol, spun down at RT and maximum speed for 5 minutes, twice. The pellet was left 30 minutes to dry and later was dissolved in a suitable volume of ddH_2O.

2.3.4.2. Phenol-Chloroform DNA extraction with phase-lock gel tubes

2.3.4.2.1. For DNA lysates

A 15 mL high-density Maxtract phase-lock gel tube QIAGEN®, was centrifuged for 1 min at 1,500xg to pellet the gel. Next, 1.6 mL of DNA lysate was poured into the tube and one equal volume of Phenol-Chloroform was added. After mixing gently by inversion, the tube was spun for 5 minutes at 1,500xg; the aqueous phase stayed on top of the gel

in the tube. Later, 2.5 volumes of 100% ethanol and 1/10 volume of 3M NaOAc pH 5.2 were added to a new 15mL tube, followed by the clear DNA supernatant recovered from the Maxtract phase-lock gel tube. The solution was mixed gently by inversion for 10 minutes until DNA fully precipitated. White DNA precipitate was then transferred to a clean 2mL tube and washed three times with 1.5 mL of 80% ethanol by inverting the tube a few times, letting it stand for 10 minutes and discarding the alcohol without centrifugation. DNA was left to air-dry completely for 1 h; next, 125 µL of TE buffer were added and kept on ice for 1 hour. Finally, the DNA was resuspended gently with a 200 µL cut tip and left for another hour on ice.

2.3.4.2.2. For digested DNA

A 2mL phase lock gel light tube was centrifuged for 1 minute at 16,000xg to pellet the gel. Next, 150 µL of digested DNA was added to the tube together with 100 µL of ddH_2O. One equal volume of Phenol-Chloroform was added and mixed gently by inversion; the tube was spun for 10 minutes at 16,000xg. Afterwards, in a clean 1.5 mL tube 2.5 volumes of 100% ethanol, 1/10 volume of 3M NaOAc pH 5.2 and 1.5 µL of Glycogen, Invitrogen™ were added and mixed by inversion with the aqueous phase containing the DNA from the phase lock gel light tube. The tube with the DNA and ethanol mix was incubated 1 h or overnight at -20°C, later it was centrifuged for 35 minutes at 16,000xg and 4°C. Supernatant was discarded and the pellet was washed once with 80% ethanol and centrifuged for 10 min at 16,000 and 4°C. Pellet was left to air-dry for 30 min; later it was resuspended carefully in 50 µL of TE buffer, and left on ice for 30 minutes; after, DNA was gently resuspended again.

2.3.5. Quantification of Nucleic Acids

Samples were quantified by placing 1 µL onto the Nano Drop column. EB buffer QIAGEN® or ddH2O were used as blank as appropriate.

2.3.6. DNA Fragment size determination by Agarose gel

To prepare a 1% agarose gel, a tray for agarose gels was placed on its base and taken into the fume hood. Next, 100mL of 1X TBE buffer were poured into a 250mL flask. 1g of Agarose was weighted and added to the buffer. The mix was heated until all the agarose was dissolved. The flask was left in the fume hood to cool down. Once it was warm, 2 µL of Ethidium Bromide were added to the solution, mixed slowly, poured onto an agarose gel tray, and a 15 well comb was placed on the top side of the gel. After, the gel solidified, the tray with the gel was placed into the tank and covered in 1X TBE buffer. samples were prepared by mixing 5 µL of DNA with 1 µL of Gel loading dye, purple 6X NEB® and loaded into the wells. 2-log ladder NEB® was used as base pair size indicator. The gel was run at 110V for 60-75 minutes.

2.3.7. Plasmid generation

2.3.7.1. Plasmid map and primers design

Plasmid maps were designed using SnapGene® software. Coding sequences for the desired protein or proteins were added to the vector, the design of the primers was carried out through the NEBuilder® Assembly Tool v1.12.17.

2.3.7.2. Polymerase Chain reaction (PCR)

Linear fragments of interest were amplified using a PCR kit supplied by Novagen®. PCR reactions and programs were set up for each specific sequence as shown on Tables 2.3 and 2.4.

	HB Domain	***HB-GFP***	***Lenti-vector***
10X KOD Hot start polymerase Buffer	5µL	5µL	5µL
dNTPs	5µL	5µL	5µL
$MgSO_4$	2µL	3µL	4.5µL

DMSO	1µL	1µL	1µL
Forward primer (10µM)	1.5µL	1.5µL	1.5µL
Reverse primer (10µM)	1.5µL	1.5µL	1.5µL
DNA Template	2µL (20ng/µL)	2µL (5ng/µL)	2µL (5ng/µL)
KOD Hot Start Polymerase	1µL	1µL	1µL
PCR grade H_20	31µL	30µL	28.5µL
Total volume	50µL	50µL	50µL

Table 2.3 PCR reactions. Reaction setup for the amplification of HB Domain, HB-GFP and Lenti-vector DNA sequences.

	HB Domain	***HB-GFP***	***Lenti-vector***
Polymerase activation	95 °C 2 min	95 °C 2 min	95 °C 2 min
Denaturation	95 °C for 20 s	95 °C for 20 s	95 °C for 20 s
Annealing	61.5°C 20 s	68°C for 20 s	59.3°C for 20 s
Elongation	68°C for 20 s	70°C for 15 s	70°C for 3:45 min
Final extension	72°C for 5 min	72°C for 5 min	72°C for 5 min
Number of cycles	35	35	35

Table 2.4 PCR programs. Thermal cycler programs used for the amplification of HB Domain, HB-GFP and Lenti-vector DNA sequences.

2.3.7.3. Ligation

Fragments of interest were ligated following the Gibson Assembly® Protocol (E5510). All assembly reactions were done using a 1:5 ratio (0.07 pmols of vector DNA and 0.35 pmols of insert DNA). Calculations for the vector and insert volumes were carried out using NEBioCalculator®; the rest of the reagent volumes are shown on Table 2.5 below.

Gibson Assembly reaction was prepared in 0.2 mL PCR tubes. Samples were incubated in the thermocycler at 50°C for 60 minutes. After, 5 µL of DNA were used to transform NEB® DH5α competent *E. coli* cells (C2987). A selection of clones was expanded, and DNA was purified using QIAGEN QIAprep® Spin Miniprep Kit. Plasmid DNA was quantified and stored at -20°C.

	Volumes
Backbone vector DNA	X µL
Insert DNA	X µL
Deionized H_20	10-X µL
Gibson Assembly Master mix (2X)	10 µL
Total Volume	20 µL

Table 2.5 Gibson assembly reaction mix. Volumes used for the ligation reaction through Gibson assembly method. Vector and insert volumes were calculated for each plasmid.

2.3.7.4. Sequencing:

After ligation, plasmid DNA was purified and quantified, samples were sent for Sanger sequencing to confirm the correct nucleotide sequence of the plasmid. Samples were prepared by mixing 5 µL of Plasmid DNA (80-100 ng/µL) and 5 µL of primer (5 µM). A total volume of 10µL was sent to GATC BIOTECH using LIGHTRUN tube barcode service. The sequences were downloaded from GATC BIOTECH website and aligned to the designed plasmid using SnapGene® software.

2.3.8. Western Blot

2.3.8.1. Cell harvest and lysis:

Protein detection by western blot started with sample collection. Media was removed, and cells were rinsed twice with ice cold PBS; Next, cells were scraped using ice cold PBS

and centrifuged for 5 minutes at 180 xg and 4°C to obtain a pellet. Alternatively, cells were rinsed once with PBS, trypsinised and centrifuged. Then the pellet was washed 2 times with ice cold PBS.

Pellets were resuspended in lysis buffer (lysis base buffer, 1 X protease inhibitor, 1X phosphatase inhibitor and Basemuncher 1:1000) and incubated on ice for 20 minutes, vortexing the samples every 5 minutes. Lysates were centrifuged at maximum speed (13200 rpm) for 20 min at 4°C and kept at -20°C.

2.3.8.2. Bradford Assay

For the quantification of proteins, a Bradford assay was performed, placing 1 mL of Coomassie Protein Assay Reagent and 1 µL of lysate into a cuvette with brief vortexing. Absorbances were read on a Jenway Genova Spectrophotometer at 595 nm.

2.3.8.3. SDS-PAGE

Separation of proteins was done through Sodium Dodecyl Sulphate-Polyacrylamide Gel Electrophoresis. Resolving gels with a concentration of 10 or 12% acrylamide were prepared as detailed on Table 2.6. To make the gel, a BIO-RAD® cassette was filled to 80% capacity with resolving gel mix, and immediately after 1 mL of isopropanol was poured into the cassette. The gel was left to polymerize. Later, isopropanol was removed, and 5 % stacking gel mix was poured into the cassette to fill the rest. Finally, a BIO-RAD® comb was inserted in the top side of the cassette and left until polymerization. Later, the comb was removed, and the cassette with the gel was placed in a BIO RAD® tank covered with 1X Running Buffer. Alternatively, 4-15% Mini-PROTEAN® TGX Stain-Free™ precast gels from BIO-RAD® were used.

Samples were prepared for loading by calculating the volume of lysate required (usually 30 to 50 µg of protein) using the values obtained from the Bradford Assay. Sample volumes were topped up with ddH_2O and 5X SDS loading buffer was added to a final concentration of 1X. Samples were mixed, boiled at 95°C for 5 minutes, centrifuged

briefly and loaded onto the gel. Precision Plus Protein Dual Colour Standard BIO-RAD® was used as a ladder. Gels were run at 120 V for 15 minutes and then at 180 V for 60 minutes. Alternatively, precast gels were run at 180 V for 45 minutes.

	Resolving gel		**Stacking gel**	
	10%	**12%**		**5%**
30 % Acrylamide	3.3 mL	4 mL	**30% Acrylamide**	0.83 mL
Tris 1.5 M (pH 8.8)	2.5 mL	2.5 mL	**Tris 1.0M (pH 6.8)**	0.63 mL
10 % SDS	0.1 mL	0.1 mL	**10% SDS**	0.05 mL
10 % APS	0.1 mL	0.1 mL	**10% APS**	0.05 mL
Ultrapure TEMED	0.004mL	0.004 mL	**Ultrapure TEMED**	0.005 mL
ddH_2O	4mL	3.3 mL	**ddH_2O**	3.4 mL
Total volume	10 mL	10 mL	**Total volume**	5 mL

Table 2.6 SDS-PAGE gel preparation. Volumes used for the preparation of 10% and 12% acrylamide gels for protein identification.

2.3.8.4. Transfer

After running the gel, proteins were transferred to a nitrocellulose membrane using a semi-dry or wet transfer methods.

2.3.8.4.1. Semi-dry transfer

On a tray, a BIO-RAD TransBlot® Turbo™ transfer filter paper stack, and a TransBlot® Turbo™ nitrocellulose membrane were soaked in 1 X transfer buffer. Next, half of the filter paper stacking was placed in the centre of a BIO-RAD TransBlot® Turbo™ machine tray, followed by the nitrocellulose membrane. After, the gel cassette was opened carefully and the gel was positioned on top of the membrane, covering it with the rest of the filter paper stacking. After each layer a roller was pressed onto the stack to make

sure all bubbles were removed. The machine was set up to Mixed or High molecular weight, to transfer all proteins of interest.

2.3.8.4.2. Wet transfer

First, wet transfer buffer was prepared (100 mL 10X Tris-Glycine transfer buffer, G-biosciences; 200 mL Methanol; 700 mL ddH_2O); next, on a tray, stacking, sponges, and the nitrocellulose membrane were soaked in this buffer. Later, on the clear side of a transfer cassette the first sponge was placed, followed by half of the stacking, the membrane, the gel; and the second half of stacking and sponge at the end. The cassette was closed and placed into a base inside of a tank with the black side facing the black side of the base. A bucket of ice and a magnetic stirrer were placed into the tank as well, and the tank was filled with wet transfer buffer. The transfer was run for 60 minutes at 400 mAmp whilst stirring.

2.3.8.5. Ponceau Staining

After transfer, the membrane was placed in a tray and covered with Ponceau S solution, SIGMA®; for 5 minutes at room temperature on a rocker; later, the membrane was washed with distilled water until the background looked white again.

2.3.8.6. Blocking

Membranes were blocked with 5% milk in TBST buffer for 1 hour at room temperature. Alternatively, XX REDACTED TEXT XX.

2.3.8.7. Antibody probing

Blocking milk was discarded after the incubation period ended, and the membrane was incubated for 1 hour at room temperature or overnight at 4°C with primary antibody diluted in 5 % milk. The working concentrations of the antibodies used are shown in Table 2.7.

Following the incubation with primary antibody, the membrane was washed 3 times with TBST for 10 minutes and the incubated with the secondary antibody (anti-mouse or anti-rabbit, BIO-RAD®; 1:4000 in 5% milk) for 1 hour at room temperature. After incubation, the membrane was washed three times with TBST buffer for 10 minutes at room temperature and later visualized on the ChemiDoc™ MP Imaging System using 1 mL of BIO-RAD Clarity™ Western ECL (Electrochemiluminescence) substrate.

To re-probe, first the membrane was washed twice with TBST buffer for 10 minutes to remove traces of ECL; then stripped with Restore™ PLUS western blot stripping buffer, Thermo Scientific™ for 15 minutes at 37°C; next, washed twice with TBST buffer for 10 minutes to remove traces of the stripping buffer, and blocked.

XX	REDACTED TEXT	XX

Antibody	Species	Company	Cat no.	Working concentration
α-Actin	Mouse	Sigma-Aldrich	A5316	1:1000
REDACTED	Rabbit	Proteintech	REDACTED	1:1000
α-DDX21	Rabbit	Novus Biologicas	NB100-1718	1:5000
REDACTED	Rabbit	Bethyl laboratories	REDACTED	1:1000
α-GFP	Rabbit	Abcam	ab6556	1:2000
REDACTED	Mouse	Abcam	REDACTED	1:1000
REDACTED	Rabbit	Bethyl laboratories	REDACTED	1:1000

REDACTED	Mouse	Cell Signaling	REDACTED	1:2000
REDACTED	Rabbit	Abcam	REDACTED	1:1000
REDACTED	Mouse	Santa Cruz	REDACTED	1:1000
REDACTED	Mouse	Cosmo Bio LTD	REDACTED	1:1000
REDACTED	Rabbit	Bethyl laboratories	REDACTED	1:1000
α-SMN	Mouse	BD	610646	1:5000
α-USP11	Rabbit	Bethyl laboratories	A-301-613A	1:2000
REDACTED	Rabbit	Bethyl laboratories	REDACTED	1:1000
α-Nucleolin	Mouse	Abcam	ab136649	1:2000
α-Nucleolin	Rabbit	Abcam	ab22758	1:2000
REDACTED		REDACTED	REDACTED	REDACTED

Table 2.7 List of primary antibodies for western blots. Species and working concentration of primary antibodies used in western blots.

2.3.9. Immunofluorescence

In the hood, coverslips were soaked in 70% IMS for a few minutes, then rinsed in ddH2O three times to remove the residues from IMS and placed in a 24-well plate to dry. In case of seeding HEK 293T cells, 400 µL of 1X poly-D-lysine was added to each well and incubated 2-16h at 37°C and 5% CO_2. After incubation, wells were washed with dH2O twice and once with PBS and left to dry inside the hood.

Around 6-8 X 10^4 cells per well, were seeded on glass coverslips in 24-well plates on day 1. When the cells reached the desired confluency, the plates were placed on ice and the media was removed from the wells. Cells were washed twice with 500µl ice-cold PBS. Next, 200µl of ice-cold methanol:acetone was added to each well to

fix/permeabilize the cells for 10 min at -20°C. Methanol:acetone was removed and cells were washed three times with 500µl of ice-cold PBS. Later, 200µl of filtered 3% BSA was added to each well and incubated at room temperature for 30min. Afterwards, BSA was aspirated and replaced by 170µl of primary antibody diluted in filtered 3% BSA as detailed on Table 2.8:

Antibody	Species	Company	Cat no.	Working concentration
α-DDX21	Rabbit	Novus Biologicals	NB100-1718	1:5000
α-GFP	Rabbit	Abcam	ab6556	1:2000
α-Nucleolin	Mouse	Abcam	ab136649	1:2000
α-Nucleolin	Rabbit	Abcam	ab22758	1:2000
S9.6 antibody (anti-R-loop)	Mouse	Kerafast	ENH001	1:500
α-SMN	Mouse	BD		1:2000
Secondary Alexa Fluor® 555	α-mouse	Thermo Fisher	a21428	1:500
Secondary Alexa Fluor® 594	α-mouse	Thermo Fisher	a11005	1:500
Secondary Alexa Fluor® 488	α-rabbit	Thermo Fisher	a11008	1:500
Secondary Alexa Fluor® 594	α-rabbit	Thermo Fisher	a110012	1:500
DAPI		Sigma	D9542	1:1000
Hoechst		Thermo Scientific		1:2000

Table 2.8 List of antibodies and dyes for immunofluorescence. Working concentrations of antibodies and dyes used in immunofluorescence.

Cells were incubated for 1h with primary antibodies; later, cells were washed with 500µl PBS three times. Next, cells were incubated with 170µl per well of secondary antibody diluted in 3%BSA with DAPI (or Hoechst) for 1 h in the dark. Finally, cells were washed with 500µl PBS three times; coverslips were rinsed in distilled water and mounted on microscope slides using Thermo Scientific Shandon™ Immu-Mount™. All slides were stored at 4°C protected from the light.

2.3.10. HB-GFP ChIP-qPCR

6-8 X 10^6 HEK293T cells were seeded in 15 cm dishes and incubated 24h at 37°C and 5% CO_2. After, cells were transfected with 10 µg plasmids EV-GFP and HB-GFP, one plasmid each dish, and incubated for 48h. Next, Cells were crosslinked by adding 37% PFA directly to the media in the dishes or flasks to a final concentration of 1%, with gentle rotation at room temperature for 10 minutes. Glycine was added to the media to a final concentration of 125mM and incubated with rotation at room temperature for 5 minutes to quench the PFA. Cells were rinsed twice with 10mL of ice-cold PBS. Later, cells were harvested by scraping with ice cold PBS, and the suspension was transferred to a 15 mL falcon tube. Cells were centrifuged for 5 minutes at 180 xg and 4°C. Pellets were resuspended in 500µL of ChIP lysis buffer and incubated for 10 minutes on ice. Alternatively, cells were trypsinised and transferred to a tube for later crosslinking and lysis as previously described.

Lysates were next sonicated; lysate volumes were split in 1.5mL Bioruptor® Pico Microtubes (250µL per tube approximately); later, tubes were placed in the Diagenode Bioruptor® Pico and cells were sonicated for 6-7 cycles (30sec ON-30sec OFF) at 4°C. After sonication, samples were centrifuged at 8,000xg and 4°C for 10 minutes and supernatants were collected into fresh 1.5 mL tubes. 100µL of sonicated sample was removed and kept aside for determination of DNA concentration, fragment size and western blot. The rest of the chromatin was kept for the immunoprecipitation (IP).

To determine the DNA concentration and fragment size, the samples were treated first with RNase A; 50µL of the lysates were topped up with ddH_2O to a final volume of 150µL,

then 2µL of RNase A (10mg/mL) was added to each sample and they were incubated at 37°C and 600rpm in the thermomixer for 30 minutes. Next to reverse the crosslinks, 6µL of 5M NaCl and 2µL of Proteinase K (20mg/mL) were added to each sample and were incubated at 65°C and 600rpm in the thermomixer for 120 minutes. DNA was purified by Phenol-Chloroform extraction as described on numeral 2.3.4.1. Fragment size was determined by running a 1% or 1.2 % agarose gel.

The immunoprecipitation was done using 50µg of chromatin and 10% of this value was set aside as input. 30µL of GFP-Trap® beads, Chromotek; were washed three times with RIPA buffer. Next, RIPA buffer was added to twice the volume of the beads and they were incubated with rotation at room temperature for 30 minutes. Meanwhile, samples were diluted 1:10 with RIPA buffer. After incubation, beads were washed once with RIPA buffer, and RIPA buffer was added to twice the volume of the beads. Next, samples were added to the beads, and incubated overnight with rotation at 4°C. The next day, beads were washed once with each of the following buffers: Low Salt wash Buffer, High salt wash Buffer and LiCl wash Buffer, in that order. Finally, 150µL of Elution buffer was added to the beads and incubated at 30°C and 800rpm for 30 minutes. Eluates were transferred to a new tube. After the IP, cross-links were reversed from eluates and inputs and DNA was purified by Phenol-Chloroform extraction as stated above, quantified, and used for qPCR.

2.3.11. γH2AX ChIP-qPCR

SMA type 1 and healthy human fibroblasts (5 T175 confluent flasks of each cell line) were crosslinked with 1% PFA. All 5 pellets per cell line were pooled and resuspended in 500µL of ChIP lysis buffer and incubated on ice for 10 minutes. Lysates were sonicated for 35 Cycles (30sec ON-30sec OFF) at 4°C in the Diagenode Bioruptor® Pico, centrifuged to pellet cell debris and the supernatants were collected on fresh 1.5 mL tubes. 10% of the lysates were set aside as input.

For the IP, 30µL of Protein G magnetic beads per sample were washed two times with PBS Tween-20 (0.02%) and incubated with 5µg of γH2AX, JBW301 cat no.05-636, Millipore®; or mouse IgG antibody, for 1 hour at room temperature with rotation. After incubation, beads were washed two times with RIPA buffer. Next, 400µL of lysates were diluted 1:1 with RIPA buffer and loaded onto the beads (400µL to the beads with the γH2AX and 400µL to the mouse IgG). Samples were incubated with the beads for 2 hours at 4°C with rotation. After incubation beads were washed four times with RIPA buffer and once with Elution buffer. Later, 150µL of Elution buffer was added to the beads and incubated at 65°C and 800rpm in the thermomixer for 30 minutes. After the IP, samples and inputs were treated with RNaseA; and Proteinase K and 5M NaCl, to reverse the cross-links. DNA was then purified by Phenol-Chloroform DNA extraction as described previously on numeral 2.3.4.1 and quantified before qPCR.

2.3.12. XX REDACTED TEXT XX

XX REDACTED TEXT XX

XX REDACTED TEXT XX

2.3.13. DRIP-qPCR

DRIP-qPCR was performed according to two different protocols with small modifications.

2.3.13.1. Modified from Ray, S., et al. 2013

Cells from six confluent T175 flasks of fibroblasts from SMA patients and healthy individuals were harvested by scraping using ice-cold PBS; 15mL tubes containing the cell suspension from each flask were centrifuged for 5 minutes at 200xg and 4°C. After removing the supernatant, the pellets were transferred to one 1.5mL Eppendorf tube per condition using 500 µL of cold PBS and centrifuged again using the same conditions. PBS was discarded and pellets resuspended in 500 µL of ChIP lysis buffer without protease inhibitor and incubated for 10 minutes on ice. Cells were sonicated 35 cycles (30 sec ON- 30 sec OFF)/ 4°C, and spun down for 10 min at 8,000xg and 4°C, to remove cell debris; supernatant was transferred to a new tube.

For the immunoprecipitation samples were divided into 3 parts of 150 µL each, labelling the 3 tubes as follows: tube1, mouse IgG; tube 2, S9.6 RNase H+; tube 3, S9.6 RNase H-. The rest of the sample was snap frozen and kept for the input. Samples were diluted 1:10 in RIPA buffer; later, 28 µg of S9.6 antibody (anti-R loop), Kerafast; or mouse IgG antibody, Thermo Fisher; were added to each tube and incubated for 1h at 4°C with rotation. Meanwhile, 30µl of Dynabeads™ protein G beads, Thermo Fisher; per tube, were washed three times with 400 µL of RIPA buffer and added to the tubes. Samples were incubated overnight at 4°C. The following day, supernatant of tube number 2 was removed, and the beads were treated with RNase H, NEB (87µL of ddH_2O, 10µL of RNase H buffer and 3 µL of RNase H) for 60 minutes at 37°C and 1000rpm. Next, supernatants were aspirated and the beads in all tubes were washed on a magnetic rack using 400 µL of Low salt wash buffer once, then High salt wash buffer and LiCl wash buffer. Tubes were centrifuged briefly at 1000 rpm to remove residual washing buffer. Elution was carried out in 150µL of Elution buffer for 15 minutes at 30°C and 1000 rpm. The eluates

containing R-loops were transferred to a fresh Eppendorf. In the meantime, 15 µL of frozen chromatin from each sample that was saved for the input was topped up to 150 µL with Elution buffer and treated with 2 µL of RNase A (10mg/mL), 2 µL of Proteinase K (20mg/mL) and 4.8 µL of 5M NaCl for 60 minutes, at 45°C and 1000 rpm. Phenol-Chloroform extraction was performed as described on 2.3.4.1 resuspending the DNA in 50 µL of PCR grade H_2O. Samples were analysed by qPCR.

2.3.13.2. Modified from Sanz L. A., Chédin, F., 2019

Cells were seeded in two 15cm plates; when cells reached 70-80% confluency, they were treated with 25 nM CPT for 10 minutes directly onto the media and incubated at 37°C and 5% CO_2. Cells were gently washed 3 times with 10ml of ice-cold PBS and harvested by scraping in ice cold PBS. Cell suspension was spun down for 5 minutes at 180 xg and 4°C. Cell pellet was resuspended in 1.6 mL of TE buffer; next, 50 µl of 20% (w/v) SDS and 10 µl of 10 mg/ml Proteinase K were added. Cell suspension was mixed by inversion 4-6 times until viscous and incubated at 4°C overnight.

Cell lysate was transferred to a previously spun 15 ml high-density Maxtract phase-lock gel, and DNA was extracted using Phenol-Chloroform as detailed on numeral 2.3.4.2 for DNA lysates. The DNA clump was cut using a 200 µL cut tip whenever it appeared to have a volume higher than 100 µL, and the enzymatic digestion was carried out into 2 separate tubes. For each digest, the enzymatic cocktail was prepared in 20 µL of RNase-free water as detailed on Table 2.9. Next, the cocktail was added to the DNA clump, followed by the rest of the water until a final volume of 150 µL was reached, and the spermidine was added at the very end to prevent DNA precipitation. Tubes were mixed by inversion 2-3 times and incubated overnight at 4°C.

Component	**Amount**	**Final concentration**
DNA	50-100 µL	
NEB buffer 2	15 µL	1X
BSA (100X)	1.5 µL	1X

Restriction enzymes EcoR1, HindIII, XbaI, Bsrg1, Ssp1	30 Units each	
RNase-free water	Up to 150 µL	
Spermidine	1.5 µL	0.5X

Table 2.9 DNA restriction enzyme digestion. Preparation of restriction enzyme cocktail for DNA digestion.

After digestion, DNA was cleaned-up using Phenol-Chloroform and 2mL high-density Maxtract phase-lock gel tubes as described on section 2.3.4.2 for digested DNA. DNA from the split sample was pooled, and 1 µL was used to quantify the concentration.

Before starting the IP, a 10µg aliquot of DNA was treated with RNase H. First, 10µL of 10X RNase H buffer was added to the DNA; next, the reaction volume was topped up to 96µL with ddH20 and 4µL of RNase H was added. The reaction mix was slightly vortexed and briefly spun down before incubation for 4-6 h at 37°C. For the immunoprecipitation, 8 µg of DNA were topped up into a final volume of 500µL with TE buffer, and a 50 µL aliquot was kept at -20°C for later use as 10% input for qPCR. Next, 52µL of 10x DRIP binding buffer and 7.14µL of S9.6 mouse antibody (2.8 µg/µL) were added to the remaining 450µL of DNA; the samples were incubated with the antibody overnight at 4°C with gentle rotation. The following day, 90µL of Protein G Dynabeads were placed into 1.5mL Eppendorf tubes. Beads were washed twice, using 700µL of 1X DRIP binding buffer for 10 minutes with gentle rotation at RT, and centrifugation for 1 minute at 1,100xg at RT. Beads were added to the samples and incubated for 2 hours at 4°C. After incubation, beads were spun down the 1 minute at 1,100 xg at RT. Next, beads were washed twice with 750µl of 1X DRIP binding buffer for 15 minutes at RT with gentle rotation, and spun down for 1 minute at 1,100 xg at RT. After washing, 300µl of DRIP elution buffer and 14µL of 10mg/mL Proteinase K were added to the beads, and the tubes were incubated at 55°C with gentle rotation for 45 minutes; later, the tubes were spun down for 1 minute at 1,100xg, and the supernatant was recovered.

R-loops were purified using 2ml high-density Maxtract phase-lock gel tubes as described on 2.3.4.2 for digested DNA, adding 300 µL of Phenol-Chloroform to the eluate (1:1) from the previous step. Samples were analysed by qPCR using the QIAGEN Rotor Gene.

2.3.14. qPCR

2.3.14.1. Rotor gene qPCR

Standard dilutions were prepared using the purified DNA 10% inputs and PCR grade H_2O in 0.2 PCR tubes, as follows; S1, (Undiluted); S2, (1:10); S3, (1:100); S4, (1:1000). The reaction mix was prepared in a 1.5 mL Eppendorf tube as detailed on Table 2.10. Later, 15µL of reaction mix was added to 0.1mL 4-strip Rotor Gene® style tubes, Starlab; and 5µL of DNA from the Standards or IP samples was added after, mixing carefully by pipetting. Alternatively, all DNA was taken to a final concentration of 2ng/µL before preparing the standards and tests as described on numeral 2.3.14.2. Tube strips were covered with caps, placed into the QIAGEN® Rotor Gene and the equipment program was set up as detailed on Table 2.11 below.

0.1mL 4-strip tubes	**Volumes 1X Reaction**
Sensimix 2X Bioline	10µL
Primer mix (5µM)	2.8µL
PCR grade H2O	2.2µL
DNA	5µL
Final volume	20µL

Table 2.10 Bioline Sensimix qPCR reaction. Preparation of reaction mix for qPCR analysis of DNA.

QIAGEN® Rotor Gene qPCR program	
Polymerase activation	95°C for 10min
Denaturation	95°C for 15s
Annealing	55°C for 15s
Extension and Acquire fluorescence	72°C for 30s
45 cycles	

Table 2.11 Rotor Gene qPCR program. Times and temperatures used for qPCR analysis

2.3.14.2. BIO-RAD® thermal cycler qPCR

Preparation of samples (inputs and IPs) was carried out by diluting them to a final concentration of 2ng/µL using qPCR grade water. Next, enough reaction mix was prepared as described on Table 2.12. Later, 8 µL of Reaction mix was placed into a well from a 96 well-plate, and 2 µL of DNA was added to the well and mixed carefully by pipetting. The plate was covered and centrifuged for 1 minute at 1760xg, then placed in the BIO-RAD® C1000 Touch Thermal cycler CFX96 Real-Time System and the program was set up as stated on Table 2.13.

96 well-plate	Volumes for 1 qPCR reaction
SYBR® green PCR master mix	5µL
Forward primer (10µM)	1µL
Reverse primer (10µM)	1µL
DNA (2ng/µL)	2µL
PCR grade H2O	1µL

Table 2.12 BIO-RAD SYBR® green qPCR reaction. Preparation of reaction mix for qPCR analysis of DNA.

BIO-RAD® qPCR program	
Initial Denaturation	95°C 5 min
Denaturation	95°C 10 sec
Annealing, extension and Acquire fluorescence	58°C 30 sec
39 cycles	

Table 2.13 BIO-RAD® qPCR program. Times and temperatures used for qPCR analysis on the BIO-RAD® C1000 Touch Thermal cycler CFX96 Real-Time System.

2.3.15. Viral production (Lentivirus)

2.3.15.1. Lentivector plasmid production

The pLenti-HB-GFP plasmid was produced through Gibson Assembly as described on numeral 2.3.7 using the pLenti-VOS plasmid as a template. Ampicillin resistant colonies were expanded to prepare 2 MIDI-preps and have enough DNA for transfection. DNA was quantified, the two tubes were mixed to have a total volume of 400 µL and the concentration was averaged.

2.3.15.2. Transfection of viral plasmids

First, 3X 10^6 HEK293T cells were seeded into 20 dishes (10cm) and incubated 24h at 37°C and 5% CO_2 in DMEM media (10% FBS, 1% Pen/Strep). Transfection mix was prepared considering 1mL of mix per dish (Final volume of 20mL): 2X HBSS was pre-warmed, and 10 mL were poured into a T25 small tissue culture (TC) flask. 5mL of 0.5M $CaCL_2$ was placed in a small TC flask. DNA from packaging plasmids (pMD.2G, 3.75µg/10cm dish; pCMVΔR8.92, 13µg/10 cm dish; pRSV-Rev 3 µg/10cm dish) and from the transfer plasmid containing the gene of interest (pLVHB-GFP 13µg/10cm dish) were combined into a small TC flask and topped up to 5mL with ddH_2O. Next, $CaCl_2$ was added to DNA dropwise. The $CaCl_2$ and DNA mix was added to the 2X HBSS dropwise, using a 5mL stripette to make bubbles for gentle mixing (solution became opalescent at pH

7.05-7.12). Finally, 1mL of the Transfection mix was added to each dish dropwise with a gentle swirl to mix. Dishes were incubated at 37°C/5%CO_2 for 6 hours (or overnight), media was changed, and dishes were further incubated for 48h.

2.3.15.3. Virus harvest

In the hood, on a paper towel previously sprayed with 70% IMS, rotor's (SW28) pots and lids were unscrewed and sprayed with IMS; next, 6 Beckman tubes were lined up and rinsed with ddH2O, same was done with the pots and lids. Two T175 flasks were placed into the cabinet; after, media from the dishes were removed with a strippette and added to one of the flasks. A 0.45 µm filter was attached to a 50mL syringe and the supernatant was filtered into the other flask.

Beckman tubes were filled with filtered supernatant to about 1.5cm below the top and placed in pots; later pots with lids were balanced with their correspondent pair. Pots were closed, placed in the rotor and into the centrifuge. Tubes were spun down at 19,000rpm for 90min at 4°C. After, the rotor was taken out the centrifuge and tubes were placed on ice. Supernatant was decanted into virkon and tubes were left inverted to remove residual media. Later, 300µL of 1% BSA in PBS was added to the pellets and left for 1h before resuspending opaque pellets. All pellets were combined and aliquoted (50µL in 0.5mL tubes). Tubes were kept at -80°C.

2.3.16. Viral titration by FACS (Lentivirus)

2.3.16.1. Transduction of HeLa cells for titration

In a 12-well plate, 75,000 HeLa cells were seeded on each well, in 1mL of EMEM (2mM Glutamine, SIGMA®; 1% NEAA, SIGMA®; 10% FBS, Gibco) and incubated overnight at 37°C/5%CO_2. Cells from one representative well were counted. Next 500µL of media was recovered from the wells about to be transduced and placed into a falcon tube. One virus aliquot was thawed on ice. Next, three viral dilutions were prepared as follows: 1:100 dilution, 10µL of virus in 990µL of conditioned media (from the falcon tube);

1:1000 dilution, 100µL of 1:100 dilution in 900µL of conditioned media; 1:10000 dilution, 100µL of 1:1000 dilution in 900µL of conditioned media.

The rest of the media from the wells to be transduced was removed and discarded; then, 450µL of the virus mix were added to the wells (two wells per virus concentration) starting with the most diluted. The rest of the wells were controls. Cells were incubated at 37°C and 5%CO_2 and after 6h wells were replenished with 500µL of fresh complete media; cells were further incubated for 72h.

2.3.16.2. Cell harvest

Medium was removed from each well and washed once with PBS, cells were trypsinised and cell suspension spun down at 500xg for 5 minutes; supernatant was discarded, pellets were resuspended in 200µL of 4% PFA and fixed for 10 minutes at RT. Cells were spun down and PFA discarded; pellets were washed once with 1 mL of PBS, centrifuged again and resuspended in 200 µL of PBS. Samples were stored at 4°C.

2.3.16.3. FACS Analysis

Transduced and controls were transferred into 1.1mL Autotubes for Flow cytometry. Samples were analysed using the BD FACS Diva 8.0.1 software set in the Blue 530-30 channel. Control samples were analysed first (non-transduced), followed by the samples with the highest GFP concentration to the most diluted. To quantify the vector titre, the data from the lowest dilution (1:10000) was used (1-20% positive cells shown as %Parent in the file) to complete the following formula:

Vector titre= [(%positive cells X no. of cells during transduction) X dilution factor X 2] = TU/mL.

2.3.16.4. MOI multiplicity of infection

On a 24-well plate with coverslips, 30,000 human fibroblasts GM08680 were seeded in 12 wells with 1 mL of media and incubated overnight 37°C and 5%CO_2. Next day, cells in a one representative well were counted (approximately 40,000 cells). Virus suspension was thawed on ice. The volume of virus needed for selected MOIs (1, 3, 10 and 30) was calculated according to the following formula:

$$\textbf{MOI} = \frac{\text{Volume of virus X titre}}{\text{Number of cells}}$$

Transduction was carried out by removing 500 µL of media from each well and placing them in 1.5mL tubes, next the volume calculated of the virus was added to the 500 µL of conditioned media. Later the remaining 500 µL of the media from the wells was discarded, and the 500 µL of viral mix was added to each well. Two wells were left untouched to use them later, as controls. After 4-6 h of incubation at 37°C and 5%CO_2, the wells were topped up to 1 mL with fresh media and the plate was incubated for 2-3 days more.

Cells were washed two times with 500 µL RT PBS and fixed with 200 µL of 4% PFA for 10 minutes at RT and then washed 3 times with PBS. Immunostaining was carried out by permeabilizing cells with 200 µL of 0.5% Triton X-100 in PBS for 10 min at RT; triton was discarded, and cells were washed 3 times with RT PBS. Next, cells were incubated in the dark for 10-20 min with 200 µL of 1:2000 Hoescht in PBS solution; Hoescht was discarded, and cells were washed 3 times with RT PBS; coverslips were mounted using IMMU-MOUNT^{TM}, Thermo scientific; mounting media. Slides were visualized on a confocal microscope to determine the percentage of cells showing GFP signal as sign of infection.

2.3.17. Lentivector HB-GFP ChIP qPCR

Primary fibroblasts from SMA patient (GMO9677) and healthy individual (GMO8680) were transduced with a lentivirus carrying HB-GFP sequence at a MOI of 30; an empty

vector-GFP virus was used as an extra control. Cells were maintained in DMEM media and 6 T175 flasks were used from each cell line for the experiment.

Cells were crosslinked at room temperature for 10 minutes with PFA (1% final concentration); reaction was quenched with Glycine (125 mM final concentration) for 5 minutes. Cells were harvested by scrapping and pellets lysed in 500 µL of ChIP lysis buffer (50 mM HEPES-KOH pH 7.5, 140 mM NaCl, 1 mM EDTA pH8, 1% Triton X-100, 0.1% Sodium Deoxycholate, 0.1% SDS, 1X Protease inhibitors) for 10 min at 4°C (on ice). Next, lysates were sonicated for 35 cycles (30sec ON-30sec OFF) at 4°C on the Diagenode Bioruptor® Pico, centrifuged at 8,000xg and 4°C for 10 minutes and supernatants were recovered. A small volume (50 µL) was used to determine DNA concentration and fragment size: crosslinks were reversed, and DNA was purified using Phenol-Chloroform, fragment sizes around 200bp were confirmed on 1% Agarose gels. The rest of the lysate was used for the immunoprecipitation (approximately 40µg of DNA per sample), samples were diluted 1:10 in RIPA buffer (50mM Tris-HCl pH 8, 150mM NaCl, 2mM EDTA pH8, 1% NP-40, 0.5% Sodium Deoxycholate, 0.1% SDS, 1X Protease inhibitors) and 30µL GFP-Trap® beads (Chromotek) were added. After overnight incubation at 4°C, beads were washed once with, Low salt wash buffer (0.1% SDS, 1% triton X-100, 2mM EDTA, 20mM Tris-HCl pH 8.0, 150mM NaCl); High salt wash buffer (0.1% SDS, 1% triton X-100, 2mM EDTA, 20mM Tris-HCl pH 8.0, 300mM NaCl); and LiCl wash buffer (0.25M LiCl, 1% NP-40, 1% Sodium Deoxycholate, 1mM EDTA, 10mM Tris-HCl pH 8.0). R-loops were eluted at 30□C and 800rpm for 30 minutes in Elution buffer (1% SDS, 100mM $NaHCO_3$). After the IP, cross-links were reversed from eluates and inputs; DNA was purified by Phenol-Chloroform extraction, and quantified. The qPCR was prepared using SensiMix™ SYBR® No-ROX kit and carried out in the Rotor Gene.

2.3.18. XX REDACTED TEXT XX

2.3.18.1. XX REDACTED TEXT XX

XX REDACTED TEXT XX

XX REDACTED TEXT XX

XX REDACTED TEXT XX

2.3.18.2. XX REDACTED TEXT XX

XX	REDACTED TEXT	XX

XX	REDACTED TEXT	XX

XX	REDACTED TEXT	XX

2.3.19. XX REDACTED TEXT XX

2.3.19.1. XX REDACTED TEXT XX

XX	REDACTED TEXT	XX

2.4. SOFTWARE ANALYSIS

2.4.1. XX REDACTED TEXT XX

XX REDACTED TEXT XX

XX REDACTED TEXT XX

2.4.2. XX REDACTED TEXT XX

XX REDACTED TEXT XX

XX REDACTED TEXT XX

2.4.3. XX REDACTED TEXT XX

XX REDACTED TEXT XX

2.4.4. Fiji software from ImageJ

Fiji from ImageJ is a software for image analysis. This tool was used for the quantification of DDX21 staining in motor neurons from mice and IPSc derived motor neurons. Images were taken in a confocal microscope. For quantification, overlapping cells, edges, and defective staining were discarded form the selections. The integrated density was the measurement considered for the image analysis.

- Opened file in FIJI (Image J). Stack viewing -> Hyperstack. Colour mode-> colorized. Ticked 'autoscale' box and click OK. Select image, clicked OK.
- To create a flat image: On the menu, clicked Image -> Stacks -> Z Project -> projection type -> Max intensity, clicked OK. The image had 2 channels: DAPI or Hoechst, and DDX21.

2.4.4.1. Nuclear signal quantification of DDX21.

2.4.4.1.1. Create binary image.

- First, with the DAPI channel active, I clicked the LUT button and selected Grays, the image changed to grayscale.
- Next, on the menu, clicked Image -> Adjust -> Threshold. Adjusted threshold values until the nuclei were covered and clicked Apply.
- On convert stack to binary pop-up window, left Method in default, background in dark and ticked the "only convert current image" box.
- A binary image was then created in the DAPI channel only. After, to make the binary image cleaner; clicked Process on the menu, then -> Binary -> Fill holes. When the software asked if I wanted to process all the images, clicked NO (the software then only processed the active channel).
- To separate cells that were thresholded together; clicked Process on the menu, then -> Binary -> Watershed. When the software asked if you want to process all the images, clicked NO.

2.4.4.1.2. Single cell quantification

- Using the Wand tracing tool, I selected the nuclear shape desired on the DAPI channel (the shape was highlighted with a yellow line); then changed to the test channel (DDX21), the shape remained selected. Clicked Analyse on the menu and then Measure. To set the measurements clicked Analyse -> Set measurements.

2.4.4.1.3. Multiple cell quantification

- Opened the ROI manager; Analyse -> Tools -> ROI manager. On the DAPI channel, after creating the binary image, selected one nuclear shape with the wand, and clicked Add in the ROI manager window (did this for all the shapes, one at a time, each nucleus should be one ROI).
- After adding all the shapes, changed channels (DDX21 channel) and clicked Measure on the **ROI manager window**. I obtained the data in the results window (If the Min and Max values are always 0 and 255 it means that the measured is for black and white channel; black is 0 and white is 255).

2.4.4.2. Nucleoplasmic signal quantification of DDX21

- After creating a binary image on the DAPI channel, used the magic wand to select the nuclear shapes and change to DDX21 channel, then threshold the image to cover the foci from the desired nuclei.
- Later holding the Alt key, selected the foci (this removed that area from the selection, resulting in the nucleoplasmic selection without the foci). Repeated this for the rest of the cells. In case that the threshold for the foci was not accurate, right click on the Oval Selection button -> Selection brush tool; then double left click on the button and a window popped up, changed pixel size to 10 (or other value as suitable). Next, with the Selection brush tool active made an accurate selection of the foci (like drawing the foci). Kept the Alt key pressed to remove a non-desired area. To add an area pressed Shift.

2.4.4.3. Foci signal quantification of DDX21

- After creating a binary image on the DAPI channel, used the magic wand to select the foci on the DDX21 channel by thresholding them (pressed Shift to select more than one foci, and Alt to remove them). If the selection was not accurate, I used the Selection Brush tool to refine the areas (pressed Shift to add an area to the selection, and Alt to remove it). Selected only the foci from one cell at a time, otherwise the values were shown as one measurement.

2.4.4.4. Export results

Measurements were obtained after clicking Analyse on the menu and then Measure. To set the measurements click Analyse -> Set measurements. Details of the regions measured were displayed in the Results window. Measurements were exported by clicking on the results window File-> Save as. The file was saved as a .csv file.

3. CHAPTER III. RIBOSOMAL DNA DAMAGE IN SURVIVAL MOTOR NEURON-DEFICIENT CELLS IS R-LOOP MEDIATED, REINFORCED BY XX REDACTED TEXT XX AND THE NUCLEOLAR DISPLACEMENT OF DEAD-BOX HELICASE DDX21.

3.1. INTRODUCTION

Spinal muscular atrophy (SMA) is an autosomal recessive motor neuron disorder, considered the main genetic cause of death in infants (Nicolau et al 2021; Wirth 2021; Cuartas and Gangwani 2022). This condition is characterised by low levels of Survival Motor Neuron (SMN) protein, generated by a homozygous deletion, or loss-of-function mutations in the *survival motor neuron 1* (*SMN1*) gene. (Prior et al., 2011; Nicolau et al., 2021). Additionally, there is a second gene, originated by a duplication event that also produces the SMN protein named *survival motor neuron 2* (*SMN2*). The severity of the disease is commonly associated with the copy number of this gene which is nearly homologous to *SMN1*. *SMN2* possesses a C to T transition and produces in most cases a truncated protein with an absent exon 7. Only 10-15% of *SMN2* transcripts produce a full-length protein which partially rescues the phenotype. Further, in absence or homozygous mutation of *SMN1,* zero copies of the *SMN2* gene is lethal; having only one copy can cause miscarriages or death by 6 months of age; meanwhile, individuals with 5 to 6 copies of the SMN2 gene do not present symptoms until adulthood (Prior et al., 2011; Schorling et al., 2020; Nicolau et al., 2021).

The SMN protein is ubiquitous and performs several different functions in the human body, including the participation in the biogenesis of small nuclear Ribonucleoproteins (snRNPs), which catalyse the removal of introns during pre-mRNA splicing; the mediation of messenger RNA (mRNA) transport in axons and modulation in the localization of β-actin in neurite growth cones, among others (Beattie and Kolb, 2018; Chaytow et al., 2018). The variety in roles of SMN, raises questions about the mechanism on which paucity of SMN protein levels result in selective degeneration of lower spinal cord motor neurons (Ahmad et al., 2016). For some time, popular hypotheses about the potential pathological mechanisms that could result in selective tissue defects have

contemplated splicing deficiencies in neuronal genes and DNA damage. These ideas were sustained by the involvement of SMN in splicing via ribosome biogenesis deficiencies. Additionally, genome stability could be compromised in post-mitotic neurons due to their high levels of transcription and the repair of the DNA lesions being mostly confined to the non-homologous end joining (NHEJ) mechanism; defects in any of these pathways could result in DNA damage (Pan et al., 2014).

Reduction of SMN correlates with a decrease in the competence of snRNPs assembly, thus affecting numerous transcripts. Nevertheless, published analysis of splicing changes in multiple genes from SMN-deficient mice showed widespread alterations present in different tissues but not changes that are specifically harmful to motor neurons (Zhang et al., 2008; Zhao et al., 2016). On the other hand, DNA damage has gained relevance in the study of SMA recently; it was first reported by Tews and Goebel in 1996 as DNA fragmentation in muscle fibres from SMA patients. More recent publications have drawn attention to the activation of p53 pathway and transcriptional R-loop accumulation/resolution as mediators of DNA damage in SMA (Jangi et al., 2017; Simon et al., 2017; Kannan et al., 2018; Kannan et al., 2020).

R-loops are nucleic acid structures where a nascent RNA hybridises with the complementary DNA strand; alterations in their metabolism have been associated with cancer and neurodegenerative diseases (Kannan et al., 2022). SMN depletion leads to R-loop accumulation and DNA damage in neuroblastoma cells (Jangi et al., 2017); additionally, SMN has been proposed to be part of an R-loop resolution pathway suggested by its interaction with Senataxin (SETX), a helicase that resolves R-loops (Zhao et al., 2016). Furthermore, overexpression of ZPR1, a zinc finger protein necessary for the accumulation and localization of SMN in cajal bodies and gems in the nucleus, reduces R-loops and increases SMN and SETX levels in SMA patient cells (Kannan et al., 2020).

Considering the aforementioned findings, the nucleolus emerged as an interesting focal point for my project. This multifunctional membraneless organelle plays an important role in coordinating the biogenesis of ribosomes. Is the residing site of ribosomal genes which are heavily transcribed by RNA polymerase I, and where the processing and modification of ribosomal RNA (rRNA) takes place. Furthermore, the high transcription rate of ribosomal DNA (rDNA), elevated content of GC regions, repetitive sequences, and replication stress, make rDNA particularly vulnerable to damage. Additionally, nucleolar stress and abnormalities have been associated to cancer and neurodegenerative disorders, such as Alzheimer's and Huntington's disease (Boisvert et al., 2007; Lindstrom et al., 2018; Xing et al., 2017; Blokhina and Buchwalter, 2020).

The aim of this Chapter was to investigate what happens in the nucleoli in SMA conditions. To determine if there are differences or defects in the biological processes carried out in the nucleolus; particularly, defects associated with rDNA transcription (such as R-loops), and if they contribute to the physiopathology of the disease as DNA damage. All this encouraged by the similarities between SMA, and other neurodegenerative diseases where DNA damage plays an important role; as well as the idea that the transcriptional process occurring inside the nucleoli in SMA, could be potentially affected as it happens in the nucleus. This analysis encompassed the assessment of nucleolar R-loop levels in SMA patient cells. The evaluation of nucleolar DNA damage using the phosphorylated H2AX as a damage marker. Additionally, I investigated the nucleolar occupancy of XX REDACTED TEXT XX XX REDACTED TEXT XX and the DEAD-box (DDX) nucleolar protein DDX21. XX REDACTED TEXT XX XX. REDACTED TEXT XX. Further, DDX21, is both involved in ribosome biogenesis and R-loop resolution.

3.2. OPTIMISATION OF METHODS FOR R-LOOP DETECTION *IN VIVO.*

3.2.1. Overview.

R-loops are DNA/RNA hybrid structures that arise during transcription and have a three-strand conformation, with the RNA bound to their complementary DNA and a displaced DNA strand. These dynamic structures are widespread and involved in several physiological mechanisms; however, non-scheduled R-loops can have pathological consequences (Sanz et al., 2016; Garcia-Muse and Aguilera 2019).

The association of co-transcriptional R-loops to neurological diseases (Groh and Gromak, 2014; Walker et al., 2017) was a source of inspiration to investigate their levels in the SMA context. In previous studies, the detection of R-loop structures have included, electron microscopy analysis (Backert, 2002); the isolated analysis of nucleic acids sensitive to ribonuclease (RNase) H that resolves DNA/RNA hybrids, and resistant to the hydrolysing effect of RNase A, an enzyme that targets single stranded RNA (Huertas and Aguilera, 2003); and the use of a catalytically inactive RNase H in Chromatin immunoprecipitation (ChIP) and Immunofluorescence (IF) experiments to trap R-loops without resolving them (Chen et al., 2017). Regardless of its sub-optimal specificity, the most common method of R-loop identification has been the S9.6 monoclonal antibody in DNA/RNA immunoprecipitation (DRIP) and immunofluorescence assays. The anti- 10 nucleotide DNA-RNA hybrid antibody S9.6, has an affinity of 0.6nM, and detects hybrid sequences as small as 6bp (Phillips et al., 2013; Garcia-Muse and Aguilera 2019). However, S9.6 also recognizes AU-rich RNA duplexes with a reduced affinity of approximately fivefold compared to DNA/RNA hybrids (Phillips et al., 2013; Halász et al., 2017). Consequently, the S9.6 method is not considered ideal for accurate detection of R-loops and is usually validated with other techniques.

Due to several issues related with the specificity of the S9.6 antibodies, for my assessment of R-loops I decided to test an alternate system, using the hybrid binding domain (HB) from the endonuclease RNase H1 fused to the green fluorescent protein (GFP) (Bhatia et al., 2014). The HB domain recognizes and binds to DNA/RNA hybrids with higher specificity for DNA/RNA heteroduplexes with a 25-fold preference to the

hybrids than double stranded RNA; and has a specificity 5-fold higher than the S9.6 antibody. (Nowotny et al., 2008). The hypothesis was that the higher specificity of the HB domain to recognize R-loops, in comparison to the S9.6 antibody, would result in more efficient binding and detection of these structures. Additionally, the presence of GFP would facilitate the tracking process in the overexpression of the fusion protein in living cells.

3.2.2. Aim and objectives.

For this Chapter, the aim was to optimise an HB-GFP-ChIP-qPCR protocol (Figure 3.1) for the detection of nucleolar R-loops in living cells. The protocol would use the HB-GFP fusion protein as a DNA/RNA-binding element to detect R-loops with higher specificity than the S9.6 antibody. Optimisation was necessary because this approach had not been tested before in our lab or in the SMA context. Potentially the protocol would later allow the identification of subtle changes in R-loop levels in the ribosomal DNA of SMA cells compared with healthy cells due to the higher specificity of the HB domain from RNAse H1.

3.2.2.1. Objectives.

- Optimise each step of the HB-GFP-ChIP-qPCR protocol in HEK293T cells.
- Select R-loop prone regions in the ribosomal DNA for qPCR analysis.
- Determine which cell lines will be used for the experiment in the SMA context.
- Optimise the protocol in the cell lines chosen in the previous step.

To fulfil this aim, according to the design of the HB-GFP-ChIP-qPCR protocol (Figure 3.1 A); first, I required to optimise the insertion and overexpression of the fusion protein HB-GFP into cells, which would bind to R-loops *in vivo.* This would be followed by paraformaldehyde crosslinking optimisation to fix the protein-R-loop bond. Next, a careful sonication testing would be necessary to shear chromatin to a size that allows the capture of R-loops through GFP immunoprecipitation. Finally, the DNA from the R-

loops would be purified and analysed via qPCR using primers targeting ribosomal DNA regions. Primers for qPCR would be selected according to their performance after testing them with the protocol. The selection of cell lines for the experiment in the SMA context, would be subjected to their availability and how well they represent SMA conditions.

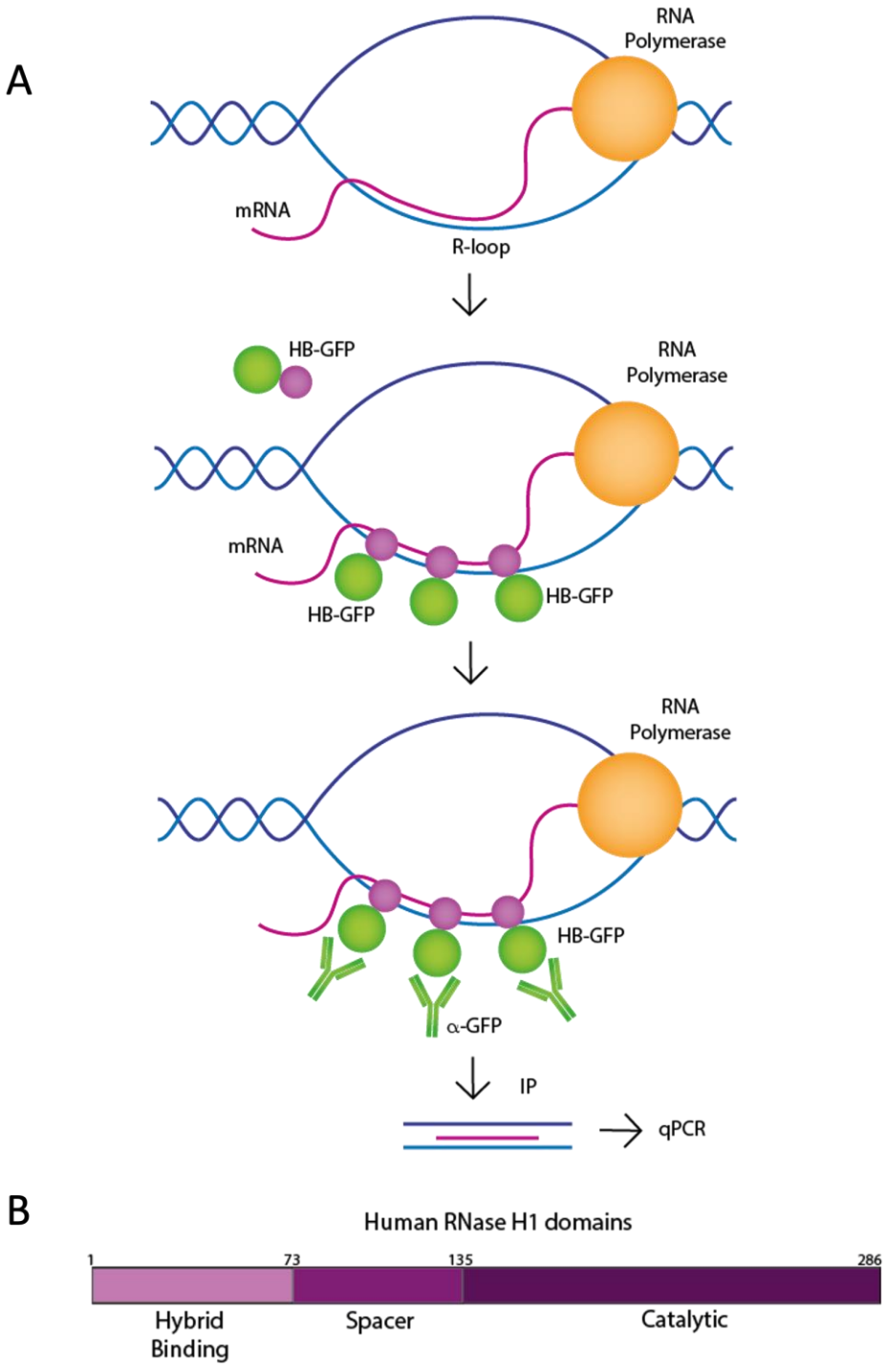

Figure 3.1. HB-GFP Chromatin immunoprecipitation approach. (A) Graphic representation of the chromatin immunoprecipitation experiment in which the fusion protein encompassed by EGFP and the Hybrid Binding domain from RNase H1, detected and bound to DNA:RNA hybrids. R-loops were isolated by pulling down the HB-GFP fusion protein with GFP-trap beads. qPCR amplified Actin 5'pause nuclear region; and 5.8S, 28S and R7 ribosomal sequences. **(B)** Domains of the human RNase H1 protein (Modified from Wu et al., 2013).

3.2.3. Results.

3.2.3.1. Optimisation of HB-GFP-ChIP-qPCR protocol in HEK293T cells.

The first step for HB-GFP-ChIP-qPCR optimisation was to transiently overexpress the HB-GFP fusion protein and the empty vector-GFP (EV-GFP) in living cells. A stable overexpression would later allow the capture of R-loops by the fusion protein. I chose HEK293T cells to test this and further steps of the protocol due to the rapid growth of the cells, and how easy they are to maintain and transfect.

HEK293T were transfected with EV-GFP or HB-GFP plasmid DNA using polyethyleneimine (PEI) to a 1:2 ratio (DNA:PEI). Protein overexpression was observed under a fluorescence microscope showing around 80% green glowing cells. Western blots probed with α-GFP antibody showed clear dark bands confirming the expression of the transfected sequences in each cell lysate. (Figure 3.2).

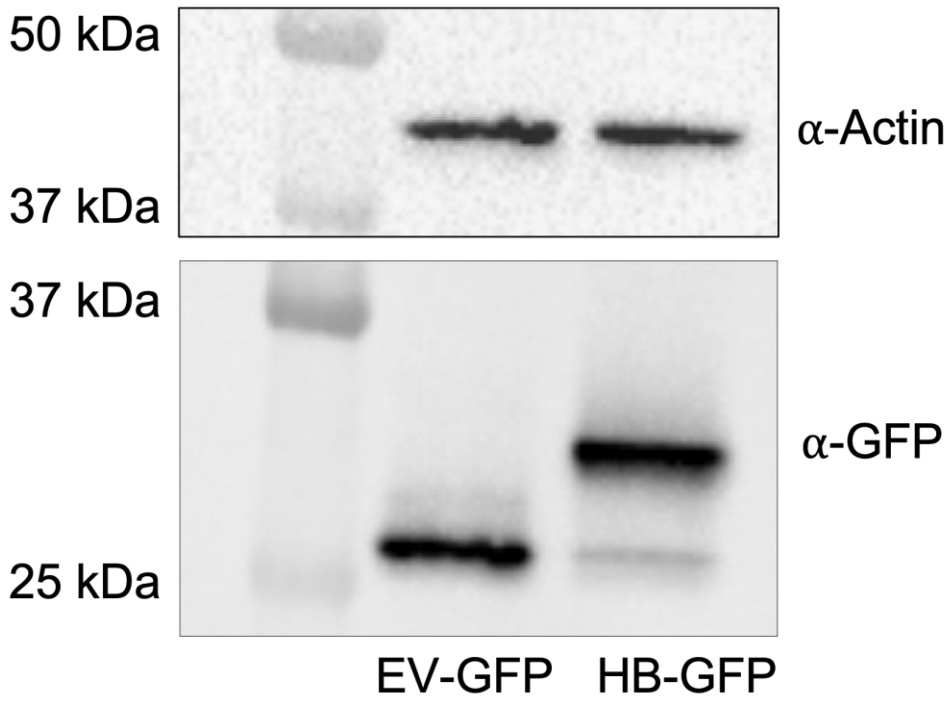

Figure 3.2. Overexpression of GFP and HB-GFP in HEK293T cells. Western blot probed with α-GFP rabbit antibody shows protein overexpression in cells transfected with EV-GFP and HB-GFP plasmids respectively. GFP band runs at 27kD and HB-GFP at approximately 33kDa.

Once the protein overexpression was consistent, samples were prepared for the immunoprecipitation. First, I started with the optimisation of paraformaldehyde (PFA) crosslinking, which maintains the interacting proteins bound to the DNA. This step is important because over-crosslinking can mask relevant epitopes recognized by the antibody during the pull-down assay (Halász et al., 2017), and hamper DNA shearing as well. Transfected HEK293T cells were cross-linked with 1% PFA for 10 minutes, and pellets were lysed in 500μL of ChIP lysis. Next, the chromatin had to be fragmented to a suitable size (approximately 200-300bp) to facilitate the pull down of specific DNA regions containing R-loops, as well as the adequate amplification of the aforementioned sequences later via qPCR. Lysates were sonicated in a Diagenode Bioruptor®, a machine that uses ultrasound waves through a cold-water bath for shearing. This technique was preferred over the restriction enzyme digestion method, since the latter results in a broader DNA size range (Halász et al., 2017). Figure 3.3 shows 1% agarose gels loaded with purified DNA from non-crosslinked (A) and crosslinked lysates (B), which usually require more cycles to reach the desired size. In both cases, 6 and 7 cycle sonication

resulted in higher uniformity of fragment size at around 200 bp with less DNA degradation, and there was no sign of over-crosslinking. Hence, the crosslinking was set at 1% PFA/10 minutes and 7 cycles of 30 sec ON - 30 sec OFF shearing.

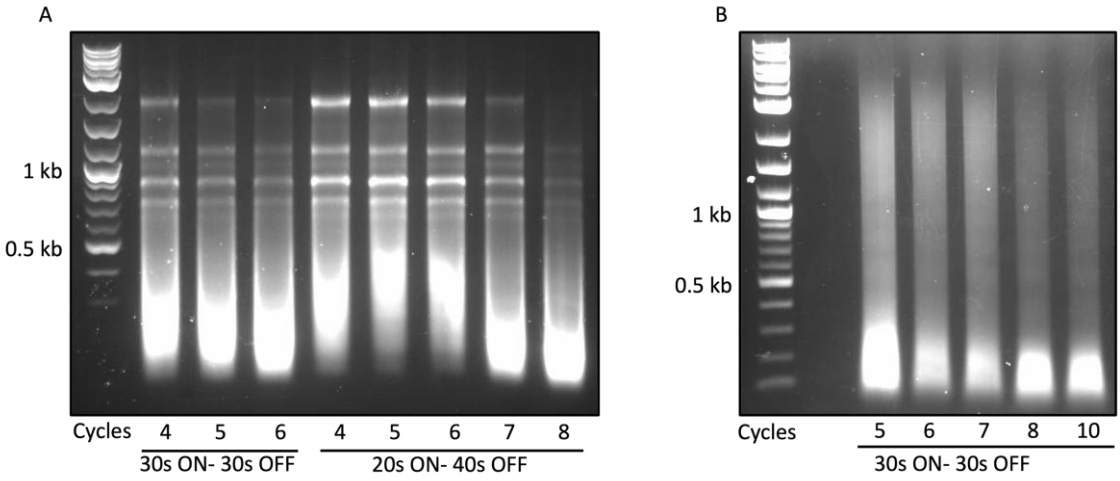

Figure 3.3. Determination of DNA fragment size. Agarose gels show DNA fragments sizes in HEK293T cell lysates during the sonication optimization process. **(A)** Six cycle sonication (30s ON-30s OFF) showed greater uniformity and adequate size for qPCR (around 200 bp) than the rest of the settings. **(B)** After crosslinking HEK293T cells, seven cycle sonication (30s ON-30s OFF) resulted in suitable DNA size without degradation.

Samples were suitable for the subsequent step; the pull-down of R-loops with α-GFP antibody. However, this was carried out using 30μL GFP-trap beads per sample instead, due to its good results in previous experiments carried out in our lab (Liao et al., 2018). The EV-GFP transfected sample was utilised as a control during the pull-down, and RNase A digestion was performed afterwards to prevent RNA contamination in all samples. Crosslinks from lysates and pulled down samples were reversed by incubating the tubes with NaCl/65°C/600rpm for 2 h, followed by the purification of DNA with the Phenol-Chloroform method.

Finally, R-loop accumulation was analysed by qPCR in the Corbett Rotor Gene-6000. Here the levels of R-loops would directly correlate with the quantification of the targeted DNA sequence. I started with a set of primers that amplify an intronic region of the Inhibitor of Growth protein 3 (ING3); a highly transcribed nuclear gene that showed a consistent elevated DRIP yield in previously published papers (Halász et al., 2017; Jurga

et al., 2021). The qPCR Ct values were evaluated with the ChIP-qPCR Data Analysis sheet from SIGMA®. The results for this experiment in the first run did not display significant difference on yields between samples; consequently, I employed as an external source of DNA damage a topoisomerase inhibitor named Camptothecin (CPT), to stall replication and favour the R-loop formation in the next experiment (Marinello et al., 2013; Jurga et al., 2021).

In the second experiment, the treatment with CPT (25μM/ 10 min) was performed immediately prior to crosslinking; the rest of the protocol was carried out with the established settings from the first experiment. Figure 3.4 presents qPCR results from the second experiment (one repeat) using ± CPT; revealing greater % Input and fold enrichment in HB-GFP transfected cells compared to EV-GFP, which becomes more apparent after CPT treatment. Furthermore, CPT does not affect the results of cells transfected with EV-GFP, confirming that only the HB-GFP fusion protein is capturing R-loops, thus confirming the suitability of the approach.

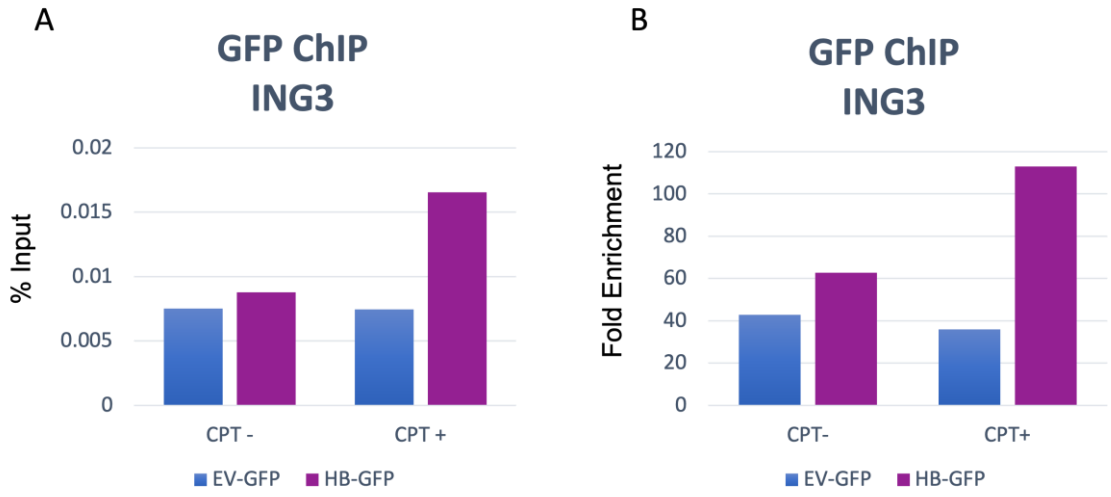

Figure 3.4. qPCR optimization using ING3 primers. Graphs show qPCR optimization results from the second experiment (one repeat) using HEK293T cells transfected with EV-GFP and HB-GFP plasmids, immunoprecipitated with α-GFP antibody. CPT treatment (25μM/10 min) was used as a validation tool to promote the accumulation of R-loops. **(A)** Plot shows higher % Input in HB-GFP transfected cells compared to EV-GFP, with a greater increase after CPT treatment. CPT does not affect results of cells transfected with EV-GFP **(B)** Fold enrichment presents the same trend.

Once the protocol for the experiment was defined, I pre-selected a variety of nuclear and ribosomal primers for R-loop prone DNA regions, within highly transcribed genes; these would probably best reflect the accumulation of R-loops in subsequent experiments. The primers included Actin 5' Pause, C-FOS, PRR5L, RPL32, SPRY2, 5.8S, 18S, 28S, R3, R5 and R7 from a few different published papers (Skourti-Stathaki et al 2011; Karahan et al., 2015, Kotsantis et al., 2016, Shen et al., 2017, Halász et al., 2017; Jurga et al., 2021). The protocol was carried out on HEK293T cells treated with CPT and I analysed only HB-GFP transfected cells after optimisation.

After running the samples through qPCR, only the primer sets that adhered to the quality controls from the Rotor Gene software were considered for further analysis. Figure 3.5 displays results from one repeat during the optimization of Actin 5' pause, RPL32 and 28S primers; graphs show higher values in % input and fold enrichment in the transduced cells when compared to the control samples. Actin 5' pause primers correspond to the 5' of 3' pause site (transcription termination site), a region prone to R-loop formation, transcribed by RNA Polymerase II (Skourti-Stathaki et al., 2011). The RPL32 gene encodes the 60S ribosomal protein L32 and is transcribed by the RNA Pol II; and the 28S primers correspond to the ribosomal DNA sequence for the ribosomal subunit 28S, transcribed by the RNA pol I. Additionally, the following primers were also chosen for this and/or subsequent experiments: 18S and 5.8S ribosomal DNA regions, as well as the R7 sequence, which comprises the end of 5.8S gene and the beginning of the internal transcribed spacer 2 (Shen et al., 2017), all are transcribed by RNA Pol I.

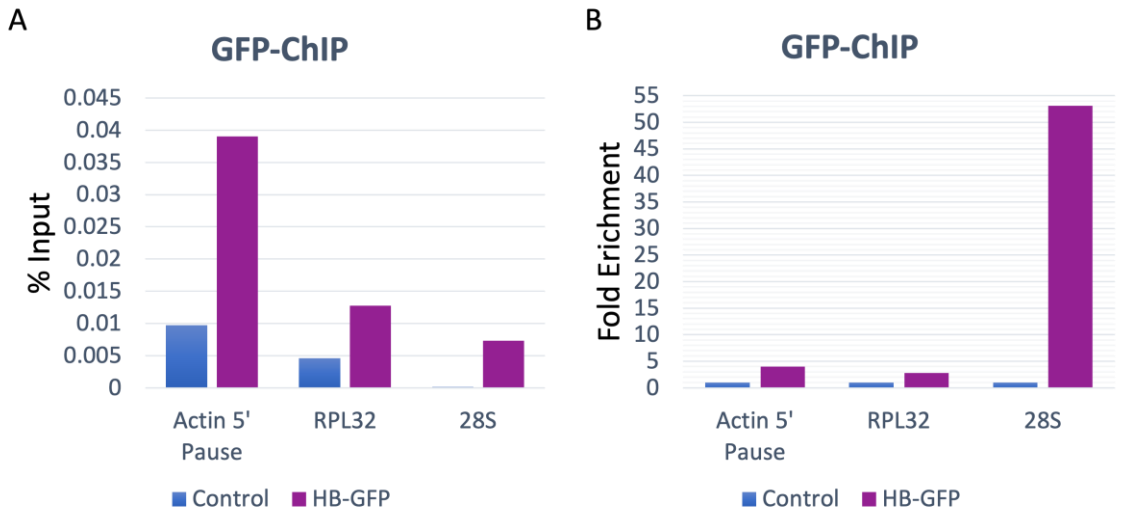

Figure 3.5. Selection of primers of interest. Plots display results from one repeat in the optimisation of primers for qPCR in HEK293T cells, transfected with HB-GFP and treated with CPT. Nuclear region Actin 5' pause, and ribosomal sequences RPL32 and 28S were selected. **(A)** Graph shows higher % Input in HB-GFP transfected cells compared to controls. **(B)** Fold Enrichment values of HB-GFP transfected cells over controls.

3.2.3.2. Optimisation of HB-GFP-ChIP-qPCR protocol in SMA context.

Considering that the main goal for the HB-GFP-ChIP-qPCR experiment is to analyse R-loop levels in the context of spinal muscular atrophy (SMA); the first objective for this second part of the optimisation was to select suitable cell lines to work with, that reflected the low survival motor neuron (SMN) protein levels present in SMA patients. An appropriate cell line selection was important because changes or extra steps would be added to the protocol for the replication of such conditions, and this could affect the feasibility of the experiment.

For the first attempt to replicate the SMA conditions, I decided to down regulate SMN protein expression in HEK293T cells. The SMN protein in human cells is encoded by the *survival motor neuron 1* (*SMN1*) gene and the highly homolog *survival motor neuron 2* (*SMN2*) gene. However, only *SMN1* produces a full-length protein in 100% of cases, since exon 7 is excluded from 90% of *SMN2* transcripts due to a silent mutation, resulting in a truncated protein (Beattie et al., 2018). Of note, more than one *SMN2* gene copy can be found in human cells, and each copy produces around 10% of full-length SMN protein

(Nicolau et al., 2021). For these reasons, the transcripts from both genes *SMN1* and *SMN2*, were targeted for silencing.

To down regulate SMN protein expression, I used two short hairpin RNA (shRNA) sequences; one targeting *SMN1* transcripts (shSMN1) and the other, *SMN2* transcripts (shSMN2). The sequences were introduced into HEK293T cells via plasmid transfection using polyethyleneimine (PEI) 1:2, DNA:PEI ratio. Additionally, I tested a vector that carried a micro-RNA (miRNA) designed to silence the *Smn* gene in rats (*Ratticus norvegicus*). According to the ensemble.com website, the *SMN1* human gene matches with 75.5% of the rat *Smn* ortholog sequence; thus, I wanted to observe if the micro-RNA would be successful at downregulating the SMN protein. The micro-RNA was inserted into the cells via plasmid transfection as well, in the same PEI ratio.

Each one of the three silencing sequences were fused with the green fluorescent protein (GFP) as a reporter gene to facilitate their visual tracking. Lysates from transfected cells were analysed by western blot and membranes were probed with α-GFP antibody to determine protein expression. Figure 3.6 shows results from the first transfection where the plasmids carrying the short hairpins shSMN1 and shSMN2 did not express adequately and had no impact over the SMN protein. Moreover, the micro-RNA was successfully overexpressed in HEK293T cells; nonetheless, SMN protein was only reduced approximately 13% in comparison with the non-transfected (NT) control, revealing that the micro-RNA was not suitable for this experiment.

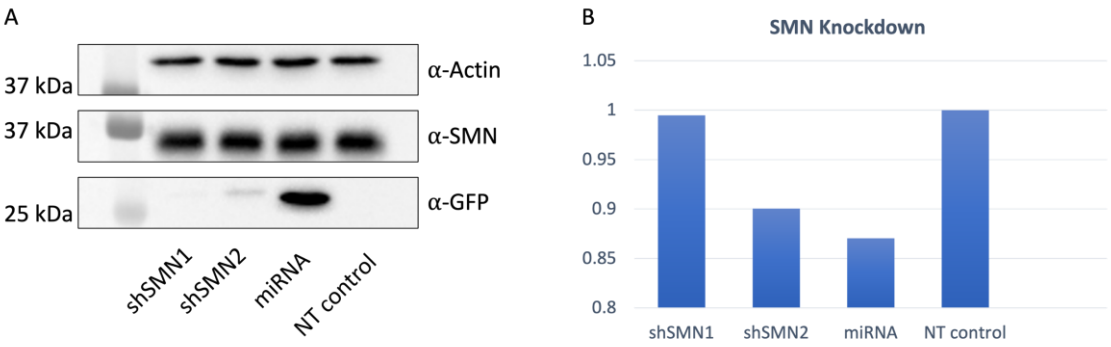

Figure 3.6. Plasmid testing for SMN knockdown on HEK293T cells. Two plasmids carried short hairpin RNA sequences to silence *SMN1* (shSMN1) or *SMN2* (shSMN2) gene transcripts. A third plasmid contained a micro-RNA (miRNA) sequence that targets the rat *Smn* gene. All sequences were coupled with GFP. shRNA sequences were hardly expressed in cells and had almost no effect on SMN expression; conversely the miRNA was strongly expressed in cells, but the knockdown was scarcely noticeable. **(A)** Western blot shows transfection efficiency as GFP overexpression, and SMN knockdown. **(B)** Graphic representation of the quantification of SMN downregulation. NT= non-transfected.

Since the shRNA sequences were not expressed properly in the first transfection, they were subjected to further optimization. During the second transfection I tested different cell density and DNA concentrations to improve the expression of the shRNAs. In 6-well plates I seeded HEK293T cells with a density of 2 x 10^5 or 3 x 10^5 and transfected them with either 1 or 2 μg of plasmid DNA, PEI ratio 1:2, DNA: PEI. After western blot analysis, I observed that in general the transfections with 2 μg of DNA resulted in an improved overexpression of the shRNAs; however, the SMN protein was not reduced further than 50% when compared with the non-transfected sample (Figure 3.7). These results were not satisfactory; therefore, my next option was to use human cells from SMA patients and healthy cells as a control.

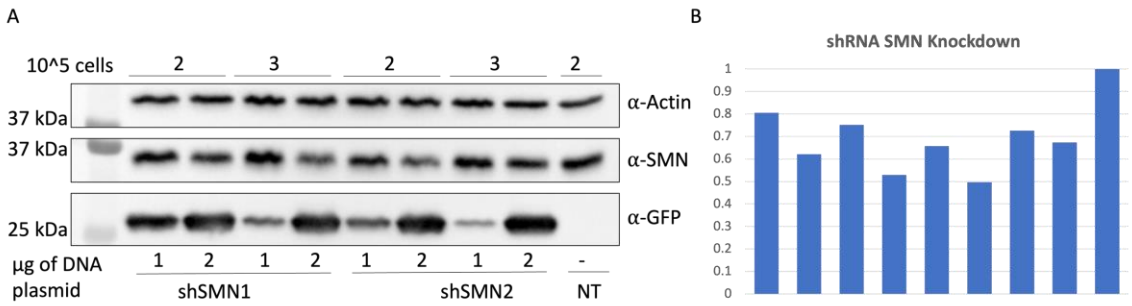

Figure 3.7. Optimization of SMN knockdown by shRNA on HEK293T cells. Optimization of the transfection using 1 or 2 μg of plasmid DNA, on wells with a cell density of 2 x 10^5 or 3 x 10^5. Transfections with 2 μg of DNA showed greater efficiency in the overexpression of the sequence; however, knockdown values were of 50% or less. NT= non transfected. **(A)** Western blot shows plasmid transfection efficiency as GFP overexpression, and SMN knockdown. **(B)** Plot displays knockdown quantification of each correspondent sample.

Spinal muscular atrophy presents in humans on different levels of severity which are classified in four types. Type I SMA is considered the most severe form, with an onset of less than 6 months old and a life expectancy of 8 to 24 months. The main determinant for SMA severity is the copy number of the *SMN2* gene, which produces about 10% of the full-length protein; type I patients possess 2-3 copies of the gene (Nicolau et al., 2021). Moreover, less severe forms of SMA present more copies of the *SMN2* gene, rescuing partially the phenotype. Therefore, cell lines with the most severe form (type I) were considered for the experiment, since they would reflect in a clearer way the molecular scene of what happens in SMA patients.

Human fibroblast cell lines were obtained from Coriell Institute repositories. The SMA type I primary cell line (GM09677) was cultured from an eye lens biopsy that belonged to a male that expired at 23 months of age as established in the type I classification. Donor presented homozygous deletion of exons 7 and 8 of the SMN1 gene, and array-based digital PCR showed 3 copies of the SMN2 gene, which match the copy number for type I SMA (Stabley et al., 2015). The healthy fibroblast cell line (GM08680) was cultured from a foreskin biopsy that belonged to an apparently healthy five-month-old male.

Briefly, in parallel with the HBGFP-ChIP-qPCR protocol, I tried to optimize a DRIP protocol modified from Ray et al. 2013, to gather more data for the overall analysis of R-loop levels. For this I used both the human fibroblasts and B-lymphocyte cells transformed with Epstein Barr virus (Coriell Institute). GM23689 lymphoblast cell line was obtained from peripheral blood from a 1-year-old affected male diagnosed with SMA type I; the healthy lymphoblast cell line AG20934, was cultured from umbilical cord blood from an apparently healthy new-born male.

For the DRIP protocol, sonicated cell lysates were divided into 3 tubes for immunoprecipitation; 28 μg of S9.6 antibody (anti-R loop), were added to two tubes and later, one of those tubes was treated with RNase H to remove R-loops (control); mouse IgG antibody Thermo Fisher was used in the last tube as a negative control. Finally, after purifying DNA from lysates and pull-down eluates, some of the previously selected

genomic regions were analysed through qPCR. Figure 3.8 presents representative results from one DRIP repeat where fold enrichment was greater in SMA fibroblasts than in healthy cells. Results were slightly inconsistent among biological repeats and needed more optimization time; however, signal from S9.6 constantly showed an elevated yield when ribosomal primers were used. This could be due to the heavy transcription carried out in the nucleoli; thus, reinforcing the decision to drive the focus of the experiment to ribosomal DNA damage.

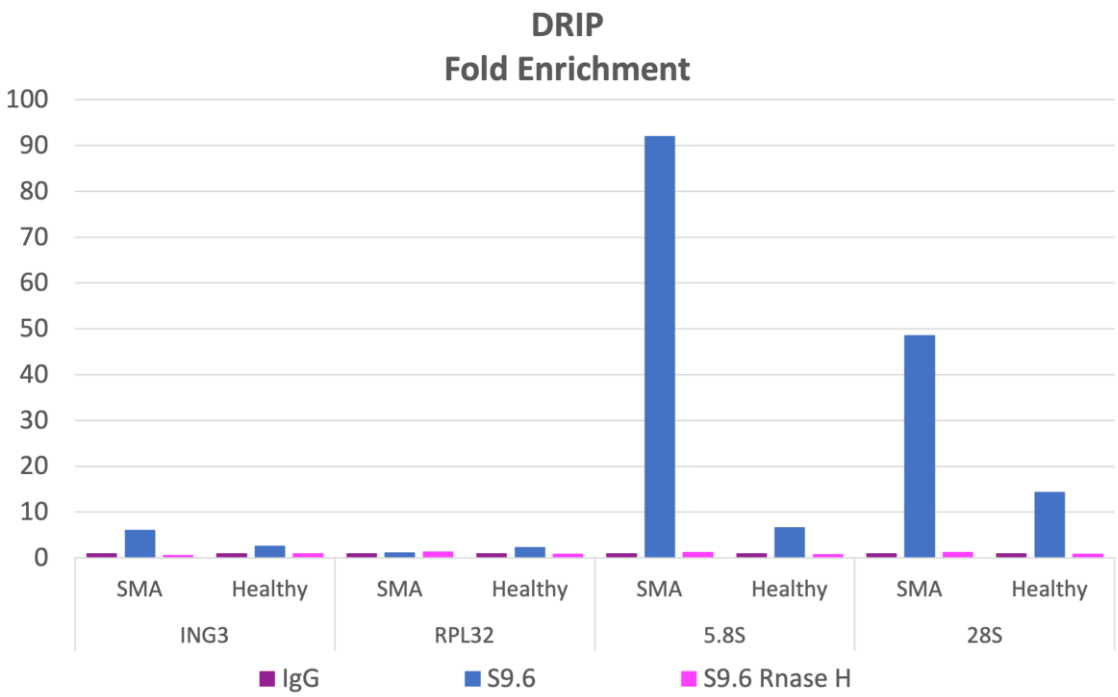

Figure 3.8. Fibroblast DRIP-qPCR optimization. Representative graph showing results from one repeat of the DRIP-qPCR optimization using human fibroblasts from SMA patients and healthy individuals. R-loop signals from the S9.6 antibody (Kerafast) were stronger in sequences amplified with ribosomal DNA primers (5.8S, 28S) than in genes transcribed by RNA Pol II (ING3, RPL32); however, results were inconsistent and further optimization was needed.

Furthermore, partial results from one repeat of B-lymphocyte DRIP (Figure 3.9) displayed greater fold enrichment in SMA lymphoblasts than in healthy cells on all tested DNA regions, supporting our hypothesis about R-loop accumulation on SMN-deficient

cells; nonetheless, values from the controls were higher than expected and needed additional optimization as well.

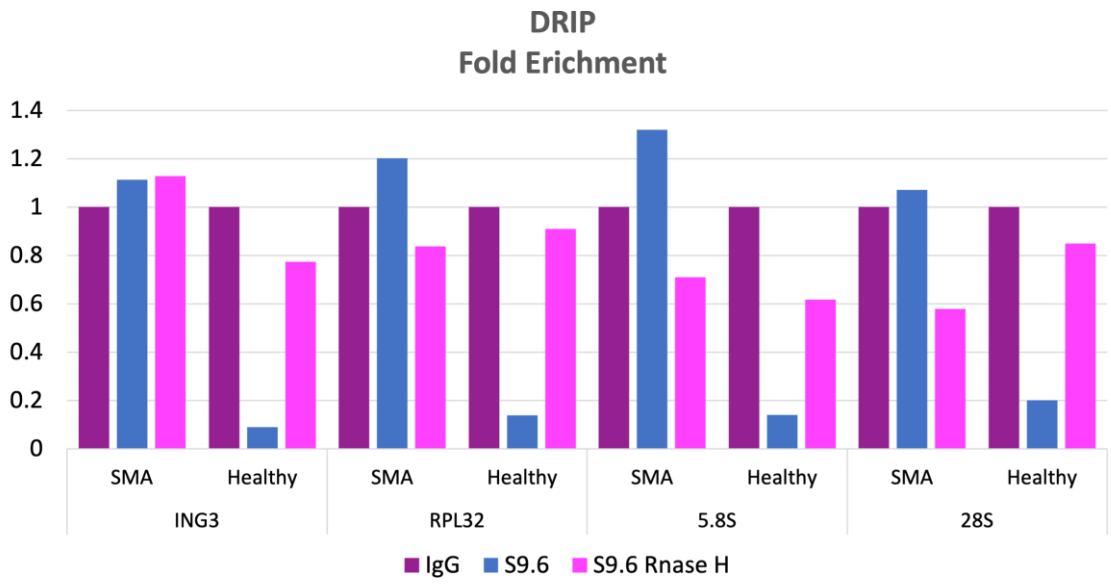

Figure 3.9. B-lymphocyte DRIP-qPCR optimization. Representative plot showing results from one repeat of the DRIP-qPCR optimization using human lymphoblasts from SMA patients and healthy individuals. R-loop signals from the S9.6 antibody were stronger in all the sequences amplified in the SMA samples (ING3, RPL32, 5.8S, 28S); however, controls needed further optimization.

Since the DRIP experiments have a sub-optimal specificity for R-loop detection and would require extra time and effort to complete the optimisation, I focused solely on the HBGFP-ChIP-qPCR protocol. As mentioned previously, the next step to continue with the HBGFP-ChIP protocol was to optimise it with cells from SMA patients and healthy individuals; the fibroblast cell lines were used for this because I found it easier to work with them during the tests for the DRIP protocol. Bearing in mind that transfection of human fibroblasts can be difficult and inefficient, a viral transduction was considered to insert the fusion protein HB-GFP into the cells, since these gene delivery vectors have shown greater efficiency (Dull et al., 1998).

In collaboration with Evangelia Karyka (Sheffield Institute of Translational Neuroscience; SITraN), we designed and produced a lentivirus that contained the genetic sequence for our fusion protein HB-GFP and a control lentivirus with GFP alone. First, the plasmid map for the lentivector containing the HB-GFP sequence was designed in SnapGene® software, and the insert was cloned into a lentiviral vector by Gibson Assembly. Later 5μL of the assembled product was transformed in DH5α competent cells (C2987, NEB®). Six single colonies were expanded, and purified plasmid DNA was sent to GATC BIOTECH for sequencing; pLVHB-GFP clone 88CD37, was selected for the rest of the experiment.

Afterwards, the production of the lentivirus (Figure 3.10) was carried out at SITraN by PhD. Evangelia Karyka and assisted by me. Twenty 10cm dishes of HEK293T cells were each transfected with a viral transfection cocktail prepared with 0.5M $CaCl_2$; 2X HBSS; pLVHB-GFP plasmid that carried our fusion protein sequence; pCMVΔR8.92, holding packaging protein sequences; pMD.2G for the envelope; and pRSV-Rev for the retro transcriptase. Virus was harvested by filtering the media from the transfected plates with 0.45μm filters (Millipore), which allow the virus to pass through but not cell debris. Filtered media was centrifuged, and pellets of virus were resuspended in 1%BSA in PBS, 50μL aliquots were kept at -80°C.

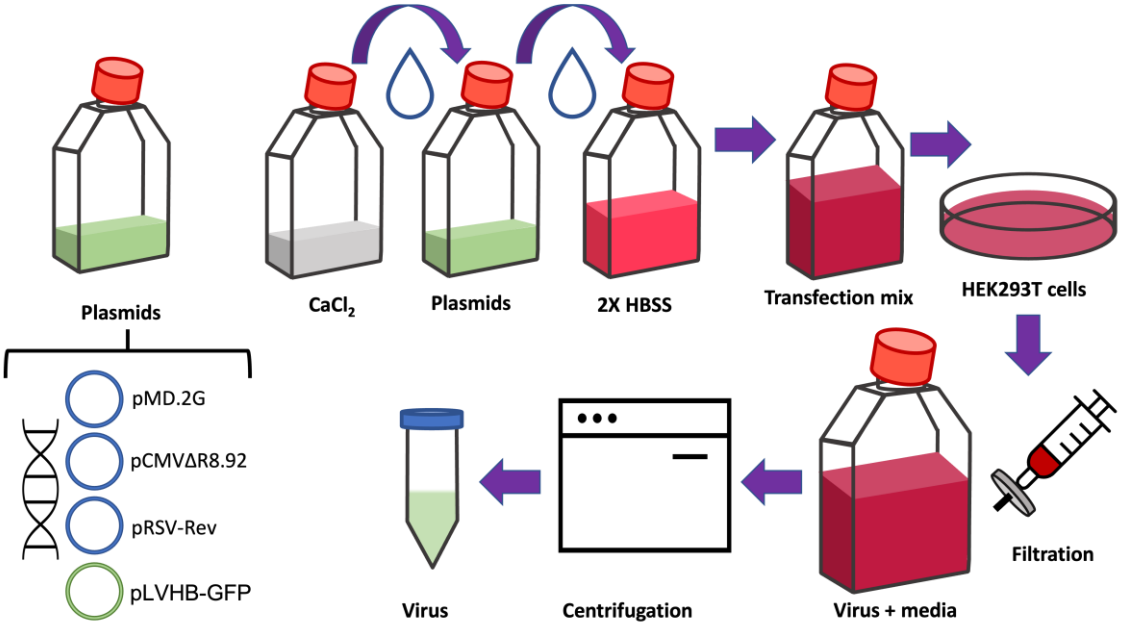

Figure 3.10. Graphic representation of virus production. Plasmids that contained the packaging sequences (pMD.2G; pCMVΔR8.92; pRSV-Rev) and the fusion gene (pLVHB-GFP) were combined in a small tissue culture flask. Next, $CaCl_2$ solution was added to the plasmid DNA dropwise. Later the DNA and $CaCl_2$ mix was added dropwise to the 2X HBSS and gently mixed; this was the transfection mix. HEK293T cells were transfected; after incubation, media was filtered to remove cell debris, leaving the virus suspended on the media. Finally, viral suspension was centrifuged, and the pellets were resuspended in 1% BSA in PBS.

Subsequently, different concentrations of the virus were transduced into HeLa cells for titration. Cells were harvested, fixed with 4% PFA and centrifuged pellets were resuspended in PBS. Detection of GFP signal by Fluorescence activated cell sorting (FACS) analysis was done at Medical School using the BD FACS Diva 8.0.1 software. Figure 3.11 presents the original results from two replicates of HeLa cell transductions with 10^-4 dilution LV-HB-GFP virus (A, B) analysed with the software; graphs on the left show the selection of cells of interest according to their size and complexity; graphs on the right present gated cells with GFP+ signal. Averaged gated cells (% Parent, GFP+) from the two replicates were used for the calculation of the vector titre as detailed below; vector titre for the HB-GFP Lentivirus was 9.5 X 10^7 TU/mL.

Vector titre= [(%positive cells x no. of cells during transduction) x dilution factor x 2] = TU/mL

LV-HB-GFP Vector titre= [(0.065 X 73,000 cells) x 10,000 x 2] = 9.5 X 10^7 TU/mL

For the control vector we followed the same protocol as for the LV-HB-GFP virus. The transfection cocktail that included the lentivector with the GFP sequence (from Professor Mimoun Azzouz; SITraN) was added to HEK293T cells; the virus was harvested and transduced into HeLa cells for titration, in duplicates. GFP signal was analysed on fixed cells by FACS. Vector titre for the control Lentivirus was 47.1 X 10^7 TU/mL.

LV-EV-GFP Vector titre= [(0.115 X 205,000 cells) x 10,000 x 2] = 47.1 X 10^7 TU/mL

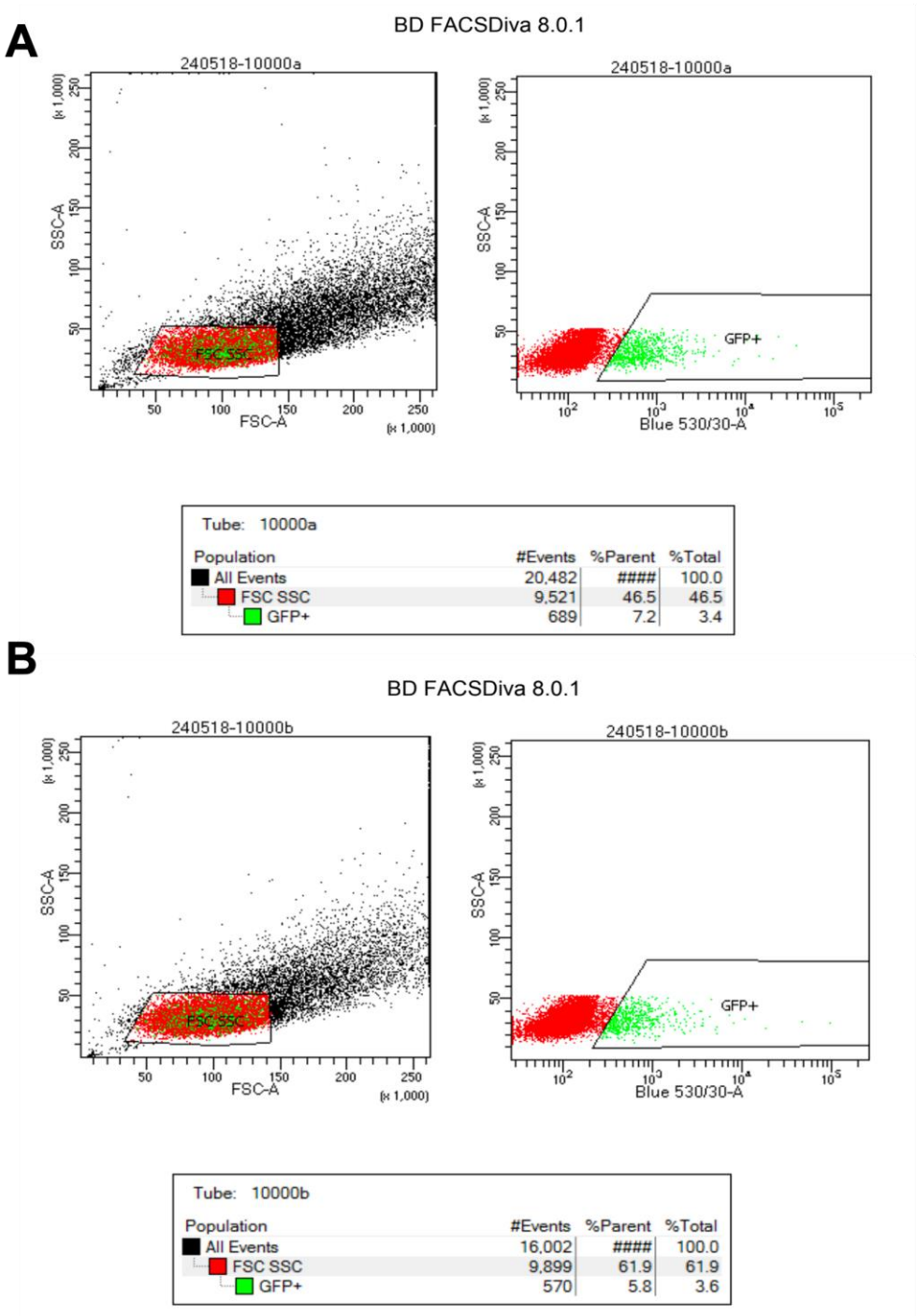

Figure 3.11. Titration of HB-GFP Lentivirus. Image from original Fluorescence activated cell sorting (FACS) results showing positively transduced HeLa cells with the HB-GFP lentivirus (10^-4 dilution). **(A)** Plot on the left presents Forward scatter (FSC; diameter of cells) against Side scatter (SSC; complexity of cells) analysis to select cells of interest. Plot on the right reveals the population of cells x 10^-4 with positive GFP fluorescence within the previously selected cells (data box, 7.2% GFP+). **(B)** Replica number two; data box shows the percentage of GFP+ cells (5.8%). Titre value calculated using GFP+ average (9.5 X 10^7 TU/mL).

The viral production was successful, and the concentrations obtained for both viruses were considered very good. Finally, the last step was to perform a Multiplicity of Infection (MOI) optimization, where we tested MOI's of 1, 3, 10 and 30. Figure 3.12 shows a confocal image from control human fibroblasts transduced with an MOI of 30, which resulted in 100% transduction efficiency. The experiment optimisation was finalised with the transduction of LV-HB-GFP or LV-GFP viruses (MOI 30), in human fibroblasts from healthy individuals and SMA patients with good results. The rest of the ChIP-qPCR protocol was subjected to a few modifications such as, number of plates/flasks and sonication cycles due to the structural differences in the fibroblast cell lines compared with HEK293T. The final protocol is described in detail on Chapter II.

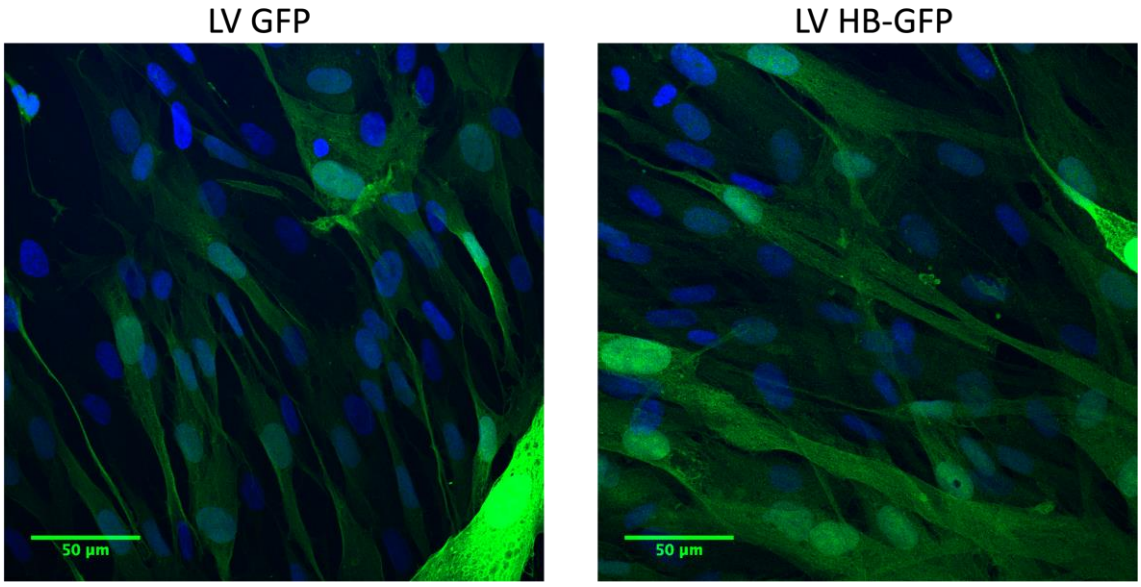

Figure 3.12. Overexpression of HB-GFP fusion protein in human fibroblasts. Confocal microscope images exhibit expression of green fluorescent protein (GFP) or HB-GFP in human fibroblasts, after transduction with LV-GFP or LV-HB-GFP viruses respectively (MOI 30). Nuclei were stained with Hoechst. Scale bar 50 µm.

3.3. ACCUMULATION OF R-LOOPS IN SMN-DEFICIENT CELLS IS ASSOCIATED WITH RIBOSOMAL DNA DAMAGE AND NUCLEOLAR DISRUPTION.

3.3.1. Overview.

For this section of the chapter, I hypothesised that it could be probable to find evidence of alterations or defects on ribosomal DNA (rDNA) in SMN-deficient cells. Particularly, the formation/accumulation of R-loop over these genomic regions; further, these hybrid structures could have damaging effects over genome stability.

Pathological R-loops can hinder physiological processes such as transcription, replication, and repair; thus, contributing to genome instability. However, in non-dividing cells, R-loop accumulation does not compromise DNA replication. The proposed pathogenic mechanisms from which R-loops contribute to the phenotypes of neurodegenerative diseases include alterations secondary to R-loop accumulation, such as, gene silencing or repeat expansion promotion, and more direct consequences like, chromatin alterations and double strand breaks, among others. (Groh et al., 2014; Garcia-Muse and Aguilera, 2019; Perego et al., 2019).

Associations between R-loops and spinal muscular atrophy (SMA) have been a subject of study over the past few years. It has been published that SMN depletion leads to intron retention, R-loop accumulation, and DNA damage in neuroblastoma cells (Jangi et al., 2017; Kannan et al., 2018), Moreover, a role in R-loop resolution has been proposed for SMN, suggested by its interaction with the helicase Senataxin (SETX) and RNA polymerase II over transcription termination sites (TTS) (Zhao et al., 2016). Although widely spread, studies have shown that R-loops tend to accumulate over highly transcribed genes, with a preference for promoters and TTS (Sanz et al., 2016; Cristini et al., 2019; Garcia-Muse and Aguilera, 2019; Jurga et al., 2021).

Ribosomal DNA is the most transcribed region in the human genome, containing hundreds of repeated genes intrinsically prone to recombination. This represents a major challenge for its replication, transcription, and repair processes (Warmerdam and

Wolthuis., 2019). Therefore, disruption on the steps involved in rRNA expression could affect the structure of nucleoli, and some of these nucleolar defects have already been found in cancers. (Schoefl, 1964; Andersen et al., 2005; Abraham et al., 2020). Recently, elevated R-loop levels in rDNA were linked to a multilobed nucleoli defect in HeLa cells (Zhou et al., 2020). Furthermore, cells from diseases associated with DNA damage such as Bloom syndrome (BLM) and Ataxia-telangiectasia (A-T) have presented signs of rDNA instability (Killen et al., 2009).

3.3.2. Aim and objectives.

The aim was to assess the presence of nucleolar R-loops in cells from SMA patients and compare them with the levels from healthy individuals. Next, to determine if there is pathogenicity associated with these structures. The reason being that rDNA is highly transcribed and R-loop prone due to its repetitive sequences, and nucleolar defects have been associated with degenerative diseases and cancers. If alterations were found, they potentially could contribute to the SMA phenotype. For this I used the optimised HB-GFP-ChIP-qPCR protocol to detect and quantify R-loops on previously selected rDNA sequences. Further, I carried out a gammaH2AX-ChIP-qPCR for DNA damage quantification.

3.3.2.1. Objectives.

- Analyse the R-loop levels in SMA and healthy fibroblasts using the optimised HB-GFP-ChIP-qPCR protocol.
- Analyse DNA damage present in SMA and healthy fibroblasts using the optimised gammaH2AX-ChIP-qPCR protocol.

3.3.3. Results.

Healthy (GM08680) and SMA type I (GM09677; 3 *SMN2* copies) fibroblasts were transduced with LV-HB-GFP virus, cells were crosslinked and sonicated for 35 cycles 30 se ON-30 sec OFF. Next, the R-loops were immunoprecipitated using GFP trap beads.

For the qPCR, I focused on nucleolar R-loops thus including three ribosomal primer sets (R7, 5.8S and 28S), and one primer set from a highly transcribed nuclear region was used as a control (Actin 5' pause). Results in Figure 3.13 although without reaching significance on most of the primer sets, present a consistent trend of higher R-loop percent input and fold enrichment values on ribosomal DNA regions from SMA fibroblasts compared to healthy cells, suggesting R-loop accumulation over ribosomal sequences. To further confirm these results, more cell lines could be added to the experiment in the future.

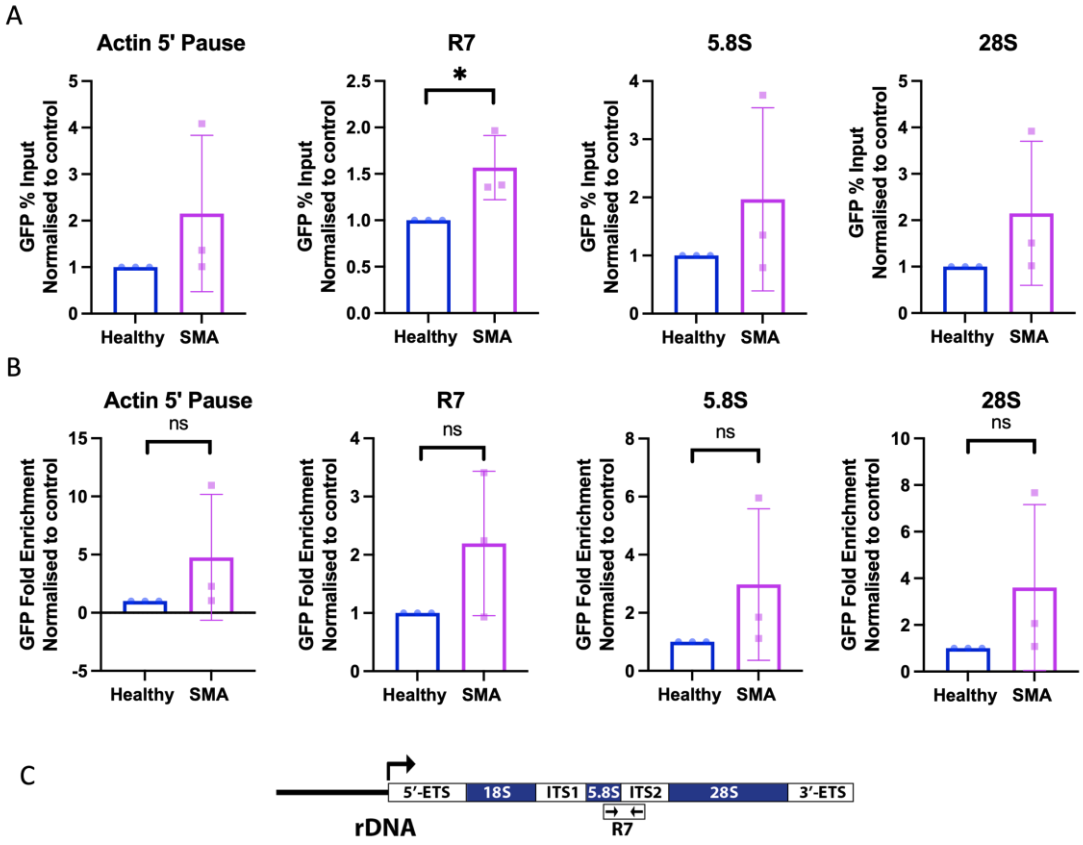

Figure 3.13. R-loops accumulate over ribosomal DNA in SMN deficient cells. Computed qPCR data from HB-GFP-ChIP experiment in SMA type I fibroblasts and healthy controls. qPCR amplified the Actin 5'pause nuclear region; and 5.8S, 28S and R7 sequences of ribosomal DNA. Data are presented as mean ± SD * $P < 0.05$; ns = not significant. The data were collected from 3 biological independent replicates (n=3). **(A)** Percent input, unpaired two tailed *t* test; p=0.3004 (Actin 5'pause), p=0.0464 (R7), p=0.3481 (5.8S), p=0.2691 (28S). **(B)** Fold enrichment, unpaired two tailed t test; p=0.2947 (Actin 5'pause), p=0.1702 (R7), p=0.2603 (5.8S), p=0.2734 (28S). **(C)** Schematic representation of the position of the R7 primers on the rDNA sequence (Modified from Shen et al., 2017).

These findings were complementary to the immunostaining experiments performed and analysed by PhD Evangelia Karyka for our paper on a different pair of fibroblast cell lines, as well as in motor neurons derived from mouse embryos (Karyka et al., 2022). Figure 3.14 A and B show SMA type I (GM03813; 3 *SMN2* copies), and control (GM00498) fibroblasts stained with S9.6 (anti-R-loop) antibody. Here, the accumulation of DNA/RNA heteroduplexes was slightly higher but not significant, following the same trend that was observed in the HB-GFP-ChIP-qPCR experiment. However, abnormal structures were observed inside of the nuclei. Additionally, both SMA type I fibroblast cell lines tested in the HB-GFP-ChIP-qPCR and immunostaining, contained 3 copies of the *SMN2* gene, which is the highest number of copies associated with SMN type I. Meaning that, it could be possible to obtain results with more significance in cell lines with fewer *SMN2* copies. Nonetheless, this would have to be tested in the future.

Later, to address the abnormal structures observed in the nuclei, fibroblasts were stained with an antibody for the major nucleolar protein, nucleolin; this made clear that the cells from SMA patients manifested significantly anomalous nucleoli, which appeared larger and clustering with each other (Figure 3.14 C and D). Furthermore, the abnormal nucleoli phenotype was visually more striking, displaying enlarged nucleoli with uneven α-nucleolin signal on SMA E13 motor neurons derived from mouse embryos when compared with the wild type (Figure 3.14 E and F). In the same way, the accumulation of R-loops in SMN-deficient (SMNΔ7) motor neurons derived from E13 murine embryos was significantly more elevated than in wild type motor neurons (Karyka et al., 2022). These results were expected since chronic low levels of SMN have a moderate impact on replicating cells but a severe effect on non-dividing cells (Cuartas and Gangwani 2022).

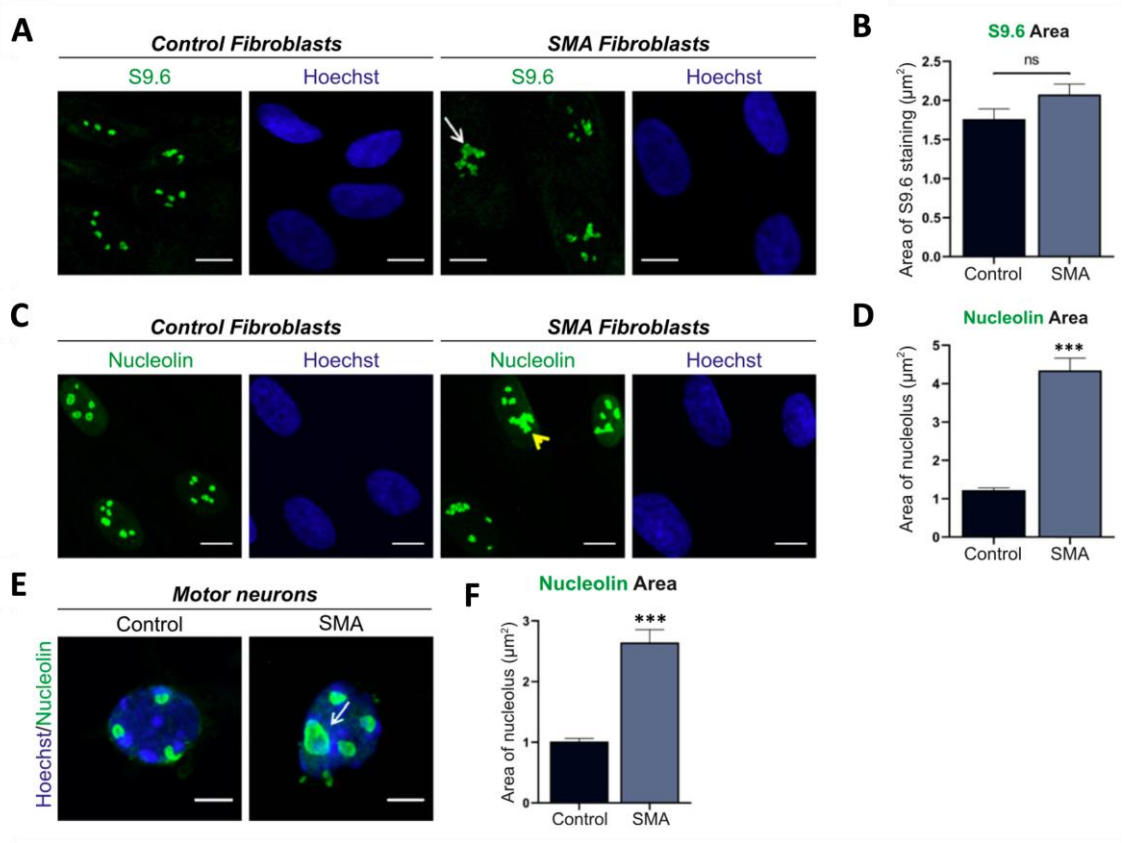

Figure 3.14. Nucleolar disruption in SMN-deficient cells. (A) R-loop (S9.6) immunostaining of fibroblasts from SMA type I patient (GM03813) and an apparently healthy individual (GM00498), display abnormal nuclear structure (arrow) on SMA cells. Scale bars 10 μm. **(B)** Quantified area (μm^2) of S9.6 immunofluorescence per cell, bars shown as mean ± SEM (n=3). Mann-Whitney non-parametric test; ns, not significant. **(C)** Nucleolin staining of SMA type I and healthy fibroblasts, exhibit abnormal nucleoli (yellow arrowhead) in SMA cells. Scale bars 10 μm. **(D)** Computed total nucleolar area per cell (nucleolin immunofluorescence) presented as mean ± SEM (n=3). Mann-Whitney non-parametric test; ***p<0.001. **(E)** Nucleolin immunofluorescence in SMA and wild type E13 embryonic murine motor neurons, reveal enlarged nucleolus (white arrow). Scale bars 5 μm. **(F)** Calculated total nucleolar area per cell shown as mean ± SEM (n=3). Mann-Whitney non-parametric test; ***p<0.001. (Partial figure from Karyka et al., 2022).

Since R-loop accumulation can result in double strand breaks (DSB), I investigated the levels of DNA damage on primary fibroblasts from SMA type I patients and healthy individuals, using the γH2AX antibody in a double immunostaining with α-nucleolin and in a chromatin immunoprecipitation-qPCR experiment. The histone variant H2AX, forms foci at sites of DNA damage, after being phosphorylated by ATM; R-loop accumulation can cause genomic instability that activates ATM (Tresini et al., 2015). The phosphorylated form of H2AX (γH2AX) has been previously used as a quantitative

indicator of double strand breaks (Rogakou et al, 1999; Modesti & Kanaar, 2001; Helena et al., 2018)

To confirm the DNA damage *in vivo*, I performed a γH2AX-ChIP-qPCR, where I sonicated human primary fibroblasts for 35 cycles (30 sec ON-30 sec OFF), and immunoprecipitated with γH2AX antibody, utilising IgG as a control. Eluates were purified by Phenol-Chloroform extraction and qPCR was performed on a BIO-RAD® C1000 Touch Thermal Cycler CFX96 Real-Time System. A ribosomal set of primers was used: RPL32, 5.8S, 18S and 28S. Figure 3.15 presents the same consistent trend observed in the HB-GFP-ChIP-qPCR experiment, with greater relative fold enrichment of γH2AX in SMN-deficient fibroblasts when compared to controls, albeit only half primer sets were significant. This suggests that the R-loop accumulation over ribosomal DNA regions could result in pathogenic double strand breaks that have a moderate impact on dividing SMA cells. Of note, this experiment was carried out with the same cell lines used in the HB-GFP-ChIP-qPCR experiment; therefore, further experimentation with additional cell lines is needed for confirmation.

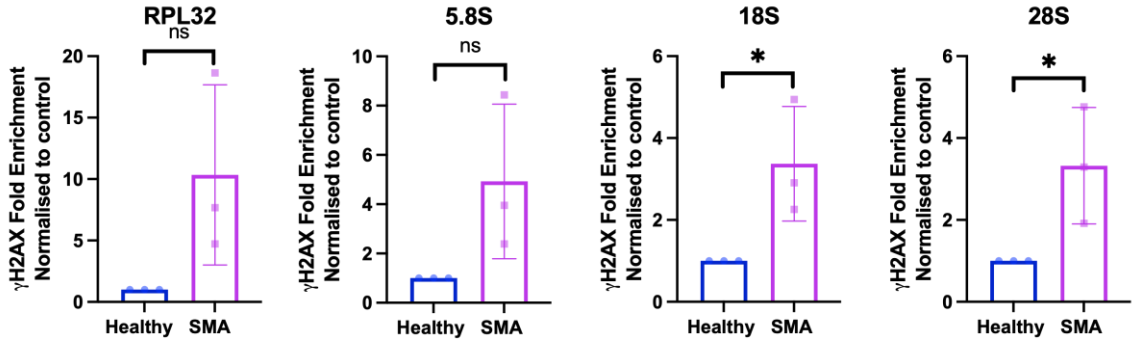

Figure 3.15. SMA cells present ribosomal DNA damage in the form of double strand breaks. **(A)** Computed qPCR data from γH2AX-ChIP experiment in SMA type I fibroblasts and healthy controls; IgG was used as a background control. qPCR amplified a region in the RPL32 gene, transcribed by RNA Pol II; and 5.8S, 18S, and 28S sequences of ribosomal DNA, transcribed by RNA Pol I. Data are presented as mean ± SD (n = 3) * $p < 0.05$; ns = not significant. Relative fold enrichment unpaired two tailed *t* test p = 0.0920 (RPL32), p = 0.0962 (5.8S), p = 0.0427 (18S), p = 0.0474 (28S).

3.4. XX REDACTED TEXT XX

3.4.1. Overview.

XX REDACTED TEXT XX

XX REDACTED TEXT XX

3.4.2. Aims and objectives.

XX REDACTED TEXT XX

3.4.2.1. Objectives.

- XX REDACTED TEXT XX.

XX REDACTED TEXT XX

3.4.3. Results

XX REDACTED TEXT XX

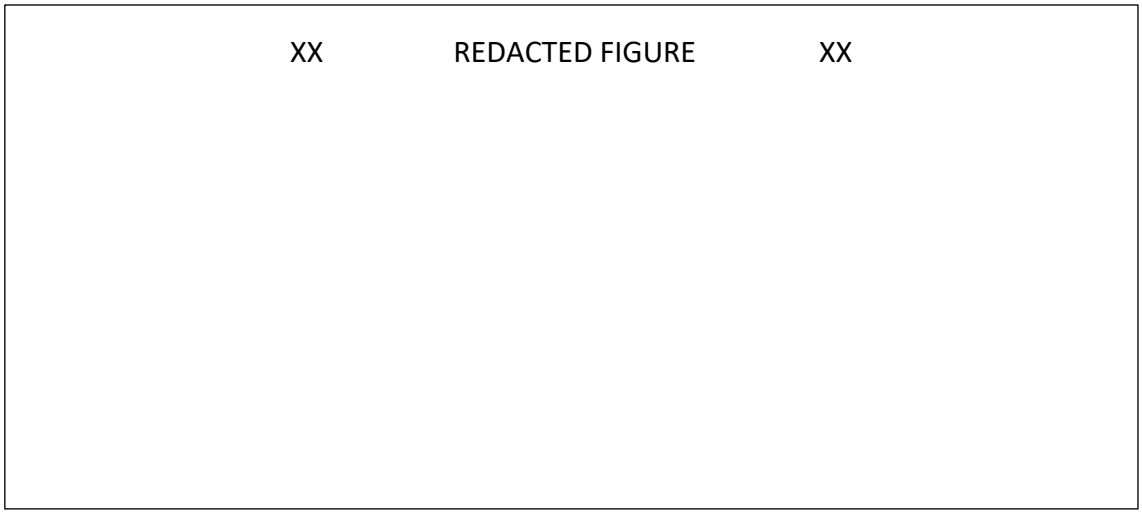

Figure 3.16. XX REDACTED TEXT XX

3.5. DDX21 IS DISPLACED FROM THE NUCLEOLUS IN SPINAL MUSCULAR ATROPHY, POTENTIALLY AS A RESULT OF R-LOOP-ASSOCIATED RIBOSOMAL DNA DAMAGE.

3.5.1. Overview.

DDX21 is a mainly nucleolar RNA helicase member of the DEAD-box (DDX) family, named after the characteristic motif, Asp-Glu-Ala-Asp (D-E-A-D), present in their amino-acid sequence. This multifunctional protein contains a highly conserved domain responsible for substrate binding, helicase, and ATPase activity (Cargill et al., 2021). DDX21 has been primarily associated with ribonucleoprotein biogenesis. This protein associates with the transcription machinery of RNA polymerases I and II, as well as with snoRNAs, to modulate ribosomal DNA (rDNA) transcription and regulate ribosomal RNA (rRNA) modification and processing (Calo et al., 2015; Sloan et al., 2015; Xing et al., 2017).

The DDX21 protein has also been linked to genome instability. In breast cancer, it is required for c-Jun phosphorylation, thus increasing the transcriptional activity of the activating protein-1 (AP-1) and promoting tumorigenesis. Moreover, ADP-ribosylation of DDX21 by active PARP-1 enhances rRNA levels and proliferation of malignant cells. (Zhang et al., 2014; Kim et al., 2019). Conversely, elevated levels of DDX21 in the early stage of colorectal cancer with microsatellite instability were found to be a positive prognosis marker for longer survival and disease-free survival (Tanaka et al., 2020).

Of particular interest for this thesis was the contribution of DDX21 to the integrity of the genome through its involvement in R-loop-associated stress, and the communication it holds with DNA damage repair (DDR) machinery. DDX21 effectively unwinds DNA/RNA hybrids. Further, deacetylation of this protein by SIRT7 increases its helicase activity improving the efficiency of R-loop resolution. Depletion of DDX21 levels or mutations on its catalytic domain causes accumulation of R-loops and increases DNA damage signal (γH2AX) in human cells (Song et al., 2017). Furthermore, induced rDNA damage in U2OS cells by I-PpoI, displaces DDX21 from the nucleoli, stimulates activation of p53 and phosphorylation of DNA-dependent protein kinase catalytic subunit (pDNA-PKcs). Moreover, inhibition of ataxia-telangiectasia mutated (ATM) or DNA-PKcs ablates the

relocation of DDX21 to the nucleoplasm in the presence of rDNA damage (Calo et al., 2018). Finally, the displacement of DDX21 from the nucleolus is one of the physiological changes implicated in craniofacial developmental defects (Calo et al., 2018), which have been found in Spinal Muscular Atrophy (Houston et al., 1994).

3.5.2. Aims and objectives.

In this part, the aim was to observe differences in the nucleolar occupancy of DDX21 in cells with low levels of survival motor neuron (SMN) protein. The purpose of this was to see if there could be an association between the R-loop and rDNA damage levels found in SMA cells, and the DDX21 levels or location. Moreover, the craniofacial developmental defects found in SMA could be the consequence of the displacement of DDX21 from the nucleoli after accumulation of rDNA damage. I assessed the occupancy of DDX21 in cells stained with α-DDX21 antibody by analysing the integrated density of the immunofluorescence using the Fiji software from Image J.

3.5.2.1. Objectives.

- To quantify the signal from the α-DDX21 antibody separately, for each compartment: nucleoli; nucleoplasm (no foci); nucleus (nucleoplasm and foci).
- Analyse the occupancy of DDX21 protein on SMNΔ7 motor neurons derived from mice embryos and controls.
- Analyse the occupancy of DDX21 protein on iPSc-derived motor neurons from SMA patient and controls.

3.5.3. Results.

To shed some light over the role that the nucleolar helicase DDX21 might play in spinal muscular atrophy (SMA). I analysed the occupancy of this protein in murine embryonic motor neurons and iPSc-derived motor neurons from SMA patients and controls that were differentiated according to the protocol published by Du et al., in 2015. The cells were stained with α-DDX21 antibody by Evangelia Karyka from SITraN.

Confocal images of p75 enriched motor neurons derived from SMNΔ7 mouse embryos showed marked differences in the intensity and localization of the DDX21 signal when compared with controls (Figure 3.17 A). Therefore, I decided to quantify three different areas; the overall fluorescence intensity in the cell nucleus, named nuclear intensity; the nucleoplasmic fluorescence intensity, which excludes the foci; and lastly the foci fluorescence intensity. The data analysis per area was carried out as follows: the fluorescence intensity from individual cells was measured; next, the values from the cells were averaged for each biological repeat. Later, the values from all the biological repeats were normalised to the general average of the control cell line. Finally, the normalised data was plotted. As presented on Figure 3.17 B and C, there was no significant difference in nuclear and nucleoplasmic fluorescence intensity between the SMA murine motor neurons and controls. However, the DDX21 signal is lower in the foci from SMA cells when compared with controls. This suggested that the DDX21 protein levels in the SMA cells were for the most part maintained, but the protein in the nucleoli was probably being displaced to the nucleoplasm.

To observe if the same phenotype presents in humans, iPSc-derived motor neurons from SMA1 patients and controls were examined. The confocal images show a reduction of DDX21 signal in the nucleoli of motor neurons derived from SMA patients compared to healthy controls (Figure 3.17 E). In this instance, a total of four cell lines were quantified: two derived from healthy individuals (csi4i, miff) and two from SMA patients (SMA86, SMA32). The analysis process was carried out in the same way as before, quantifying the nuclear, nucleoplasmic and foci fluorescence intensity of DDX21. Later, the average values from each biological repeat were normalised using the general average value from both control samples. Data from each cell line was plotted individually and combined (control vs SMA). Figure 3.17 F and G, show no significant difference in nuclear fluorescence intensity between the cell lines individually or combined. Likewise, Figure 3.17 H and I, reveal similar nucleoplasmic fluorescence intensity values among the cell lines individually and combined. Nevertheless, the fluorescence intensity of the foci was visually lower in the SMA cell lines compared to the controls (Figure 3.17 J). Further, the

difference became significant when the cell lines were combined (Figure 3.16 K). This suggests that DDX21 is possibly migrating to the nucleoplasm in human cells as well.

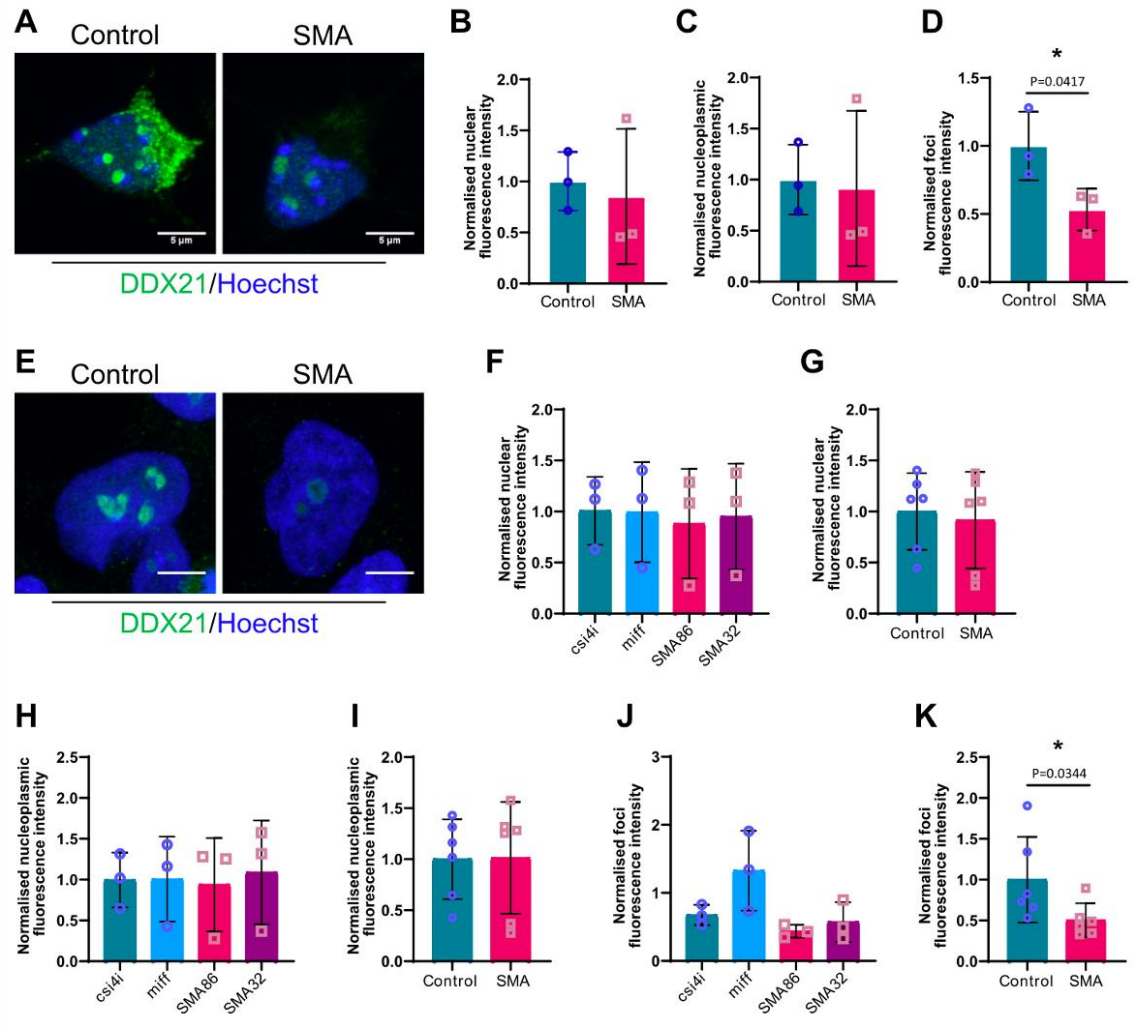

Figure 3.17. SMN-deficient motor neurons display reduction in nucleolar DDX21 signal compared with controls. (A) DDX21 staining of p75 enriched motor neurons (DIV7) derived from SMNΔ7 and control mouse embryos. Scale bars 5µm. **(B)** Nuclear DDX21 Fluorescence intensity normalised to control **(C)** Nucleoplasmic intensity (excluding foci) **(D)** Foci intensity. Data presented as mean ± S.D. (n=3). *p < 0.05; paired two tailed *t* test (p=0.0417). **(E)** DDX21 labelling of iPSc-derived motor neurons isolated from SMA type I patients (SMA86, SMA32) and healthy controls (csi4i, miff). Scale bars 5µm. **(F)** Graph of DDX21 nuclear fluorescence intensity normalised against average of control samples. **(G)** Plot displaying combined nuclear intensity from controls (csi4i, miff) and SMA (SMA68, SMA32) samples. **(H)** DDX21 Nucleoplasmic fluorescence intensity (excluding foci) normalised against average of control samples. **(I)** Combined nuclear intensity from controls (csi4i, miff) and SMA (SMA68, SMA32) samples. **(J)** Graph of DDX21 foci fluorescence intensity normalised against average of control samples. **(K)** Combined foci intensity from controls (csi4i, miff) and SMA (SMA68, SMA32) samples. Data presented as mean ± S.D. (n=3). *p < 0.05; paired two tailed *t* test (p=0.0344). (From Karyka et al., 2022)

Recapitulation of chapter data.

- R-loops accumulate over ribosomal DNA, in cells from spinal muscular atrophy (SMA) patient.
- Immunofluorescence examination with S9.6 and α-nucleolin antibodies, revealed disruption in the nucleolar structure of survival motor neuron (SMN)-deficient cells.
- ChIP-qPCR analysis using gammaH2AX showed that SMA patient cells accumulate DNA damage over ribosomal gene regions.
- XX REDACTED TEXT XX

XX REDACTED TEXT XX

XX REDACTED TEXT XX

- Immunofluorescence analysis with α-DDX21 antibody revealed that in SMN-deficient cells DDX21 is displaced from the nucleoli to the nucleoplasm.

3.6. DISCUSSION.

3.6.1. HB-GFP-ChIP-qPCR as a method for R-loop detection in vivo.

R-loops have both beneficial and detrimental roles. These three-stranded nucleic acid structures can behave as intermediates that facilitate some cellular processes. However, they can also become a source of DNA damage when unscheduled (Garcia-Muse and Aguilera, 2019). This dichotomy creates the need for accurate detection of these structures, whose distribution over the human genome varies under different conditions. The development and optimisation of reliable methods to detect DNA/RNA hybrids is still evolving. Previously R-loop detection mainly relied on the use of S9.6 antibody; however lately the use of a catalytically inactive RNaseH1 has become more prominent.

Here I present a method for R-loop detection, adapted for the specific needs that the study of spinal muscular atrophy required. The protocol uses the hybrid binding (HB) domain from the RNAseH1 endonuclease, fused to the green fluorescent protein (GFP). The fusion protein binds to R-loops *in vivo*, preventing their resolution. Later, hybrids are recovered by immunoprecipitation with GFP-trap beads and purified DNA is analysed by qPCR. This method allowed the detection of changes in the levels of nucleolar R-loops in survival motor neuron (SMN)-deficient cells when compared with controls.

The process of protocol optimisation for the analysis of structures such as R-loops is generally complex, with multiple steps and challenges. Methods based on S9.6 generally have a low signal-to-noise ratio due to antibody specificity issues; and R-loops are usually enriched over gene bodies and transcription termination sites (TTS) (Smolka et al., 2021; Wang et al., 2021). Conversely, RNaseH1 based approaches have shown better resolution, but most R-loops detected are located around promoters and transcription start sites (TSS) (Chen et al., 2017; Yan et al., 2019). This is probably the result of the catalytically inactive RNAseH1 binding to the R-loop but not resolving it; thus, preventing the continuation of transcription.

For the protocol described here (HB-GFP-ChIP-qPCR), I used only the hybrid binding domain from RNaseH1 fused to GFP, instead of the catalytically inactive full protein to detect DNA/RNA hybrids. The purpose of this was to have a smaller size fusion protein to ease its overexpression in human cells and favour its access to the nucleoli. Once developed a lentivirus as a delivery system, the stable overexpression of the protein in human fibroblasts was successful. Nevertheless, over time, GFP degradation was observed after passage 8 post-transfection; therefore, only early passages were used for the analysis. In general, the protocol is very straight-forward and less time consuming than other methods. The presence of GFP in the fusion protein facilitated the visualisation of its overexpression in the cell lines. Further, this method allowed the identification of R-loop accumulation on fibroblasts, a cell type that is not typically considered affected in SMA. When compared with other approaches, the main disadvantages for this protocol were the overexpression of the fusion protein in cells, along with monitoring its degradation; the preference of HB for transcription start sites and promoters; and that this method does not identify R-loops in a strand specific fashion.

Currently, there are new techniques that try to address the limitations on R-loop mapping. BisMapR, is a technique that uses a catalytically inactive RNAseH1 fused with micrococcal nuclease (MNase) from MapR. (Yan et al., 2019). Next, sodium bisulfite is used on the released R-loops to deaminate the cytosines into uracils on the displaced DNA strand for later degradation. Later, it provides strand specific identification of R-loops (Wulfridge and Sarma, 2021). Another interesting approach is the R-loop CUT&Tag. This method uses an Tn5-based cleavage under target and tagmentation (Lu et al., 2020), and the GST-His_6-2XHBD fusion protein (HBD: hybrid binding domain from RNaseH1) as a sensor to recognise the hybrids. The identification of R-loops through this technique has better resolution than DRIPc-seq, and more sensitivity than MapR and R-ChIP in the detection of R-loops located over gene bodies. Nonetheless, it does not provide strand specific localization of the loops. Furthermore, replacement of GST-His_6-2XHBD with S9.6 antibody in R-loop CUT&Tag revealed similar results. (Wang et al., 2021). Finally, at this moment in time, efforts to produce more suitable methods for R-

loop detection are focusing more on improving the means for capture of DNA/RNA hybrids. Steps in protocols that could be a source of noise or cause the loss of fragile R-loops such as sonication, fixation, restriction digestion or creation of stable cell lines, are slowly being replaced for less invasive/aggressive approaches as seen in MapR and R-loop CUT&Tag.

3.6.2. R-loops as a source of ribosomal DNA damage in Spinal muscular atrophy.

Accumulation of R-loops and DNA damage in spinal muscular atrophy (SMA), has been studied previously (Jangi et al., 2017; Kannan et al., 2018). However, the possibility that part of this accumulation happens over ribosomal DNA (rDNA) in SMA has not been addressed before. Here, I used the optimised HB-GFP-ChIP-qPCR protocol for R-loop detection, and γH2AX-ChIP-qPCR for double strand breaks (DSB). These techniques were carried out on fibroblasts from patients with severe spinal muscular atrophy (GM09677; SMA type 1). Results revealed a trend in the accumulation of R-loops over rDNA in that SMA cell line (Figure 3.12). Further, that subtle accumulation resulted in a trend of increased DNA damage on these regions (Figure 3.14). Moreover, it is important to consider further testing ribosomal R-loop and γH2AX levels through ChIP-qPCR on other fibroblast cell lines to confirm these findings; perhaps on a SMA cell line with less copies of the *SMN2* gene.

The elevated number of R-loops was no surprise due to the high transcription rates of rDNA, which represent around half of the overall cellular transcription (Moss et al., 2002). Moreover, the GC content in rDNA is greater than the genome-wide average, elevating the probability of R-loop formation (Escobar et al., 2011). Conversely, DNA/RNA hybrids are also purposely set over intergenic spacers of rDNA by the RNA polymerase II, to avert production of sense intergenic noncoding RNAs (sincRNAs) by RNA Polymerase I; thus, preventing defects on rRNA biogenesis and disruption of nucleolar structure. (Abraham et al., 2020). Any imbalance on the components involved in R-loop homeostasis or a combination could incite their dysregulation.

Immunofluorescence analysis with S9.6 antibody on SMA fibroblasts published in our paper (Karyka et al., 2022), showed the fluorescent signal localising in clusters that resembled abnormal nucleoli. Further single staining of the fibroblasts with the nucleolar antibody α-nucleolin, revealed the same anomalous foci pattern, suggesting that the structure of nucleoli in SMA cells was aberrant (Figure 3.13 A and C). However, to confirm that the R-loops were indeed co-localising mainly with nucleoli would require double staining and further analysis on dividing patient cell lines and motor neurons (iPSc/murine). Moreover, sincRNAs levels could be addressed to determine if there is any compensation through R-loop accumulation in these SMA cell lines.

Interestingly, these results were obtained from a cell line that is not typically considered vulnerable to the chronic effects of low survival motor neuron (SMN) protein levels. Similar observations were published by Kannan et al., 2018, where immunostaining signals from S9.6 antibody (R-loops) and γH2AX accumulated in the nucleus of fibroblasts from SMA patients. Further, DNA-PKcs were downregulated in dividing cells from SMA patients, possibly hindering the non-homologous end joining (NHEJ) repair pathway (Kannan et al., 2018). Here, my results support that chronic low SMN levels have a moderate effect on what happens in the nucleus and nucleoli of replicating cells.

The relevance of R-loop and DNA damage accumulation on the nuclear and ribosomal DNA from replicating and non-dividing cells resides on the suggestions made over the past decade by a number of researchers, about SMA having a systemic physiological impact. This has been supported by clinical reports and murine model research, which have revealed multiple non-neuromuscular phenotypes in SMA patients that include a variety of systems affected, such as the gastrointestinal, skeletal, reproductive, metabolic, and cardiovascular, as well as the detection of similar defects in mice models. (Zhang et al., 2008; Shababi et al., 2014; Lipnick et al., 2019; Kim et al., 2020).

A network of collaborating proteins associated with R-loop resolution and DNA damage repair have been found altered in SMA. ZPR1, a zinc finger protein that upregulates SMN

expression, is downregulated in SMA. This protein forms complexes with the DNA/RNA helicase senataxin (SETX), SMN and R-loops. Deficiencies of ZPR1 or SMN downregulate SETX and DNA-dependent protein kinases (DNA-PKcs). Further, overexpression of SMN, SETX or ZPR1, reduces R-loops and rescues the SMA phenotype (Zhao et al., 2016; Kannan et al., 2020; Kannan et al., 2022). DNA-PKcs are an important element of the NHEJ, the primary DNA repair pathway for neurons. Interestingly, although there is evidence of rDNA DSB being resolved by homologous recombination (HR) and NHEJ. Inhibition of the NHEJ pathway showed an increase in rDNA DSB, as well as transcriptional inhibition. Whereas inhibition of HR had no impact on DSBs or obstructed cell cycle progression; suggesting that rDNA DSB could be preferably processed through the NHEJ pathway (Harding et al., 2015; Blokhina and Buchwalter, 2020).

3.6.3. R-loop driven nucleolar disruption in SMN-deficient cells.

The nucleolus possesses a three-layered architecture that reflects the stages of ribosome biogenesis and depends on ribosomal DNA (rDNA) transcription status (van Sluis and McStay, 2017). As mentioned in the previous section, the accumulation of R-loops over rDNA, results in double strand breaks (DBS). Immunofluorescence experiments presented in this thesis provide evidence of abnormalities in the nucleolar structure of SMN-deficient cells, as a result of rDNA damage. SMA Fibroblasts stained with the nucleolar antibody α-nucleolin revealed aberrant nucleolar structures clumping together (Figure 3.13 C). Further analysis with α-nucleolin on motor neurons derived from murine embryos showed enlarged nucleoli in SMA cells when compared with controls (Figure 3.13 E). Moreover, in our paper (Karyka et al., 2022), double staining of SMA fibroblasts with α-nucleolin, exhibited several γH2AX foci co-localizing with cap-like structures in the periphery of the nucleoli.

The re-organization of the rDNA into forming nucleolar caps occurs after the inhibition of RNA Polymerase I (Pol I) transcription. In the presence of ribosomal DNA damage, transcription is inhibited by ataxia-telangiectasia mutated (ATM). Later, rDNA DSB migrate to the periphery of the nucleoli to be repaired. Nowadays, it is believed that

NHEJ acts mostly within the nucleoli, rapidly resolving DSB; and HR to be active at nucleolar caps, where persistent breaks might relocate. Ku70/80 a protein that promotes NHEJ has been detected inside the nucleoli, whereas HR factors mainly concentrate at the caps. Further, signal from γH2AX is generally found next to nucleolar caps, possibly due to the lack of nucleosome occupancy over transcriptionally active rDNA inside the nucleoli (Harding et al., 2015; van Sluis and McStay 2017; Warmerdam and Wolthuis, 2019; Blokhina and Buchwalter, 2020).

In this case, the nucleolar abnormalities observed, follow the described response to rDNA DSB. Additionally, the reported downregulation of DNA-PKcs in SMN-deficient cells by Kannan et al., could be contributing to the nucleolar phenotype, since depletion of DNA-PKcs leads to an increase of rDNA DSB and cap formation (Blokhina and Buchwalter, 2020; Kannan et al., 2020).

3.6.4. Displacement of DEAD-box helicase DDX21 from nucleoli in SMN-deficient cells and its role in R-loop resolution.

Analysis through immunofluorescence carried out on motor neurons derived from mice embryos as well as human iPSc-derived motor neurons in this thesis, showed that in SMN-deficient cells the DDX21 DNA/RNA helicase is displaced from the nucleoli into the nucleoplasm. DDX21 main roles involve the regulation of ribosome biogenesis and ribosomal RNA (rRNA) processing. However, the helicase function of DDX21 is required for proper binding when interacting with other molecules. (Calo et al., 2015; Song et al., 2017).

In my results the rDNA damage generated by R-loop accumulation in SMN-deficient cells could be triggering relocation of DDX21 from the nucleoli into the nucleoplasm. DDX21 has been reported to be displaced from the nucleolus after rDNA damage induced by the I-PpoI endonuclease on U2OS cells; after damage linked to treacle ribosome biogenesis factor 1 (TCOF1) protein dysfunction; and upon stabilization of the transcription factor p53 (Calo et al., 2018). Moreover, in the presence of rDNA damage,

local transcription ceases in the nucleoli, and the rDNA is rearranged to facilitate repair; interestingly, inhibition of the repair factors ATM or DNA-PKcs, prevents migration of DDX21 to the nucleoplasm, even in the presence of rDNA damage (Calo et al., 2018). Suggesting that what drives DDX21 to migrate could be associated with DNA damage signalling.

In SMN-deficient cells displacement of the DNA/RNA helicase DDX21 could be triggered by the R-loop-driven rDNA damage. Meaning that, the presence of DDX21 in the nucleoli could be conditioned by active transcription; particularly, through its possible participation in the resolution of nucleolar R-loops. This protein interacts with the Pol I transcription complex in the nucleoli; and has been reported to cooperate in the resolution of genomic R-loops in breast cancer cells (Calo et al., 2015; Song et al., 2017). Further, DDX21 has been found to be highly expressed in the nucleoli in some cancers, promoting rDNA transcription and cell proliferation (Zhang et al., 2014; Kim et al., 2019). This supports the importance of DDX21 during transcription in the nucleoli and how it could become dispensable when the process becomes stalled. However, it is still unclear if DDX21 migrates from the nucleoli slowly upon R-loop accumulation on rDNA due to disturbances in rDNA transcription progression caused by low SMN levels and its interaction with RNA polymerase I (Karyka et al., 2022); thus, hindering further the resolution of nucleolar DNA/RNA hybrids. Or if dysregulation of other factors such as DNA damage signalling proteins (ATM, DNA-PK) are affecting its nucleolar localization, and that is generating accumulation of R-loops on SMN-deficient cells (Figure 3.18).

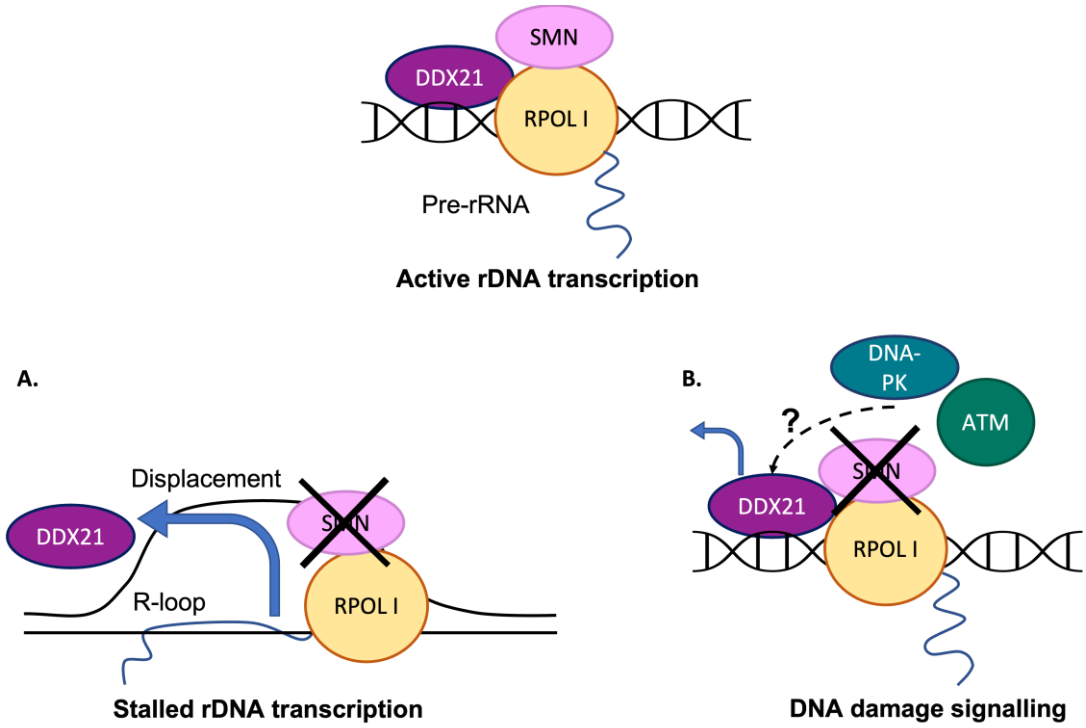

Figure 3.18. Hypothesis for DDX21 nucleolar displacement in SMN-deficient cells. DDX21 in the nucleoli has been associated with active rDNA transcription. In SMN-deficient cells **(A)** DDX21 could be displaced due to R-loop formation caused by low levels of SMN; or **(B)** Dysregulation of DNA damage signalling factors by low SMN levels could be affecting DDX21 nucleolar occupation.

3.6.5. XX. REDACTED TEXT XX.

XX REDACTED TEXT XX

XX REDACTED TEXT XX

4. CHAPTER IV. XX REDACTED TEXT XX.

4.1. INTRODUCTION

XX REDACTED TEXT XX

XX REDACTED TEXT XX

XX REDACTED TEXT XX

XX REDACTED TEXT XX

XX REDACTED TEXT XX

XX REDACTED TEXT XX

XX REDACTED TEXT XX

4.2. XX REDACTED TEXT XX.

4.2.1. Overview

XX REDACTED TEXT XX

XX REDACTED TEXT XX

XX REDACTED TEXT XX

4.2.2. Aims and objectives.

XX REDACTED TEXT XX

4.2.2.1. Objectives

XX REDACTED TEXT XX

4.2.3. Results

XX REDACTED TEXT XX

XX REDACTED FIGURE XX

Figure 4.1. XX REDACTED TEXT XX

XX REDACTED TEXT XX

XX REDACTED FIGURE XX

Figure 4.2. XX REDACTED TEXT XX

XX REDACTED TEXT XX

XX REDACTED FIGURE XX

Figure 4.3. XX REDACTED TEXT XX

XX REDACTED TEXT XX

XX REDACTED TEXT XX

XX REDACTED TEXT XX

XX REDACTED FIGURE XX

Figure 4.4. XX REDACTED TEXT XX.

XX REDACTED TEXT XX

XX REDACTED FIGURE XX

Figure 4.5. XX REDACTED TEXT XX

XX REDACTED TEXT XX

XX REDACTED FIGURE XX

Figure 4.6. XX REDACTED TEXT XX.

XX REDACTED FIGURE XX

Figure 4.7. XX REDACTED TEXT XX.

XX REDACTED TEXT XX

XX REDACTED FIGURE XX

Figure 4.8. XX REDACTED TEXT XX.

XX REDACTED FIGURE XX

Figure 4.9. XX REDACTED TEXT XX.

XX REDACTED TEXT XX

XX REDACTED FIGURE XX

Figure 4.10. XX REDACTED TEXT XX.

XX REDACTED TEXT XX

XX REDACTED TEXT XX

XX REDACTED TABLE XX

Table 4.1. XX REDACTED TEXT XX.

XX REDACTED TEXT XX

XX REDACTED TEXT XX

Figure 4.11. XX REDACTED TEXT XX.

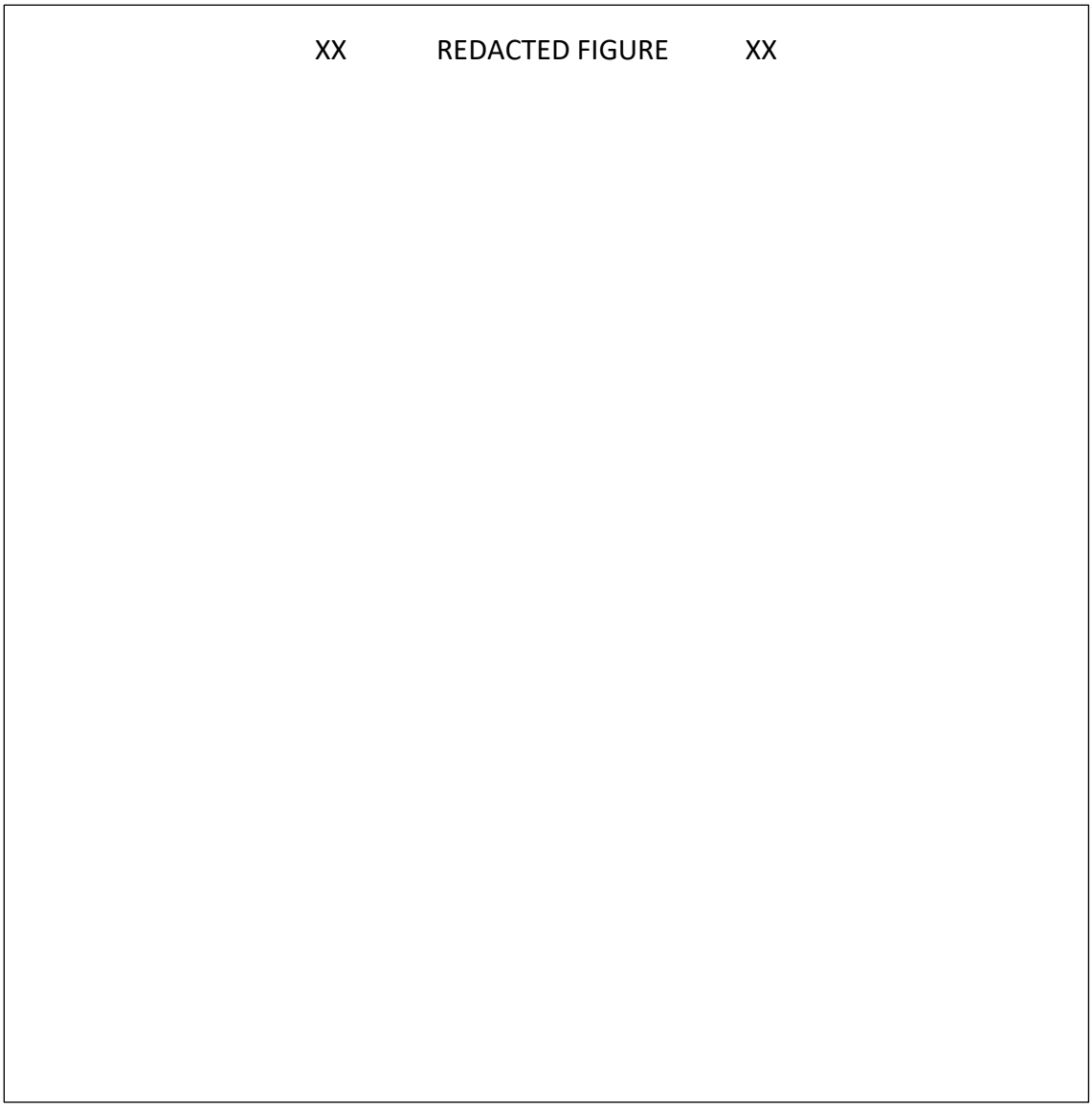

Figure 4.12. XX REDACTED TEXT XX.

XX REDACTED TEXT XX

XX REDACTED TEXT XX

XX REDACTED FIGURE XX

Figure 4.13. XX REDACTED TEXT XX.

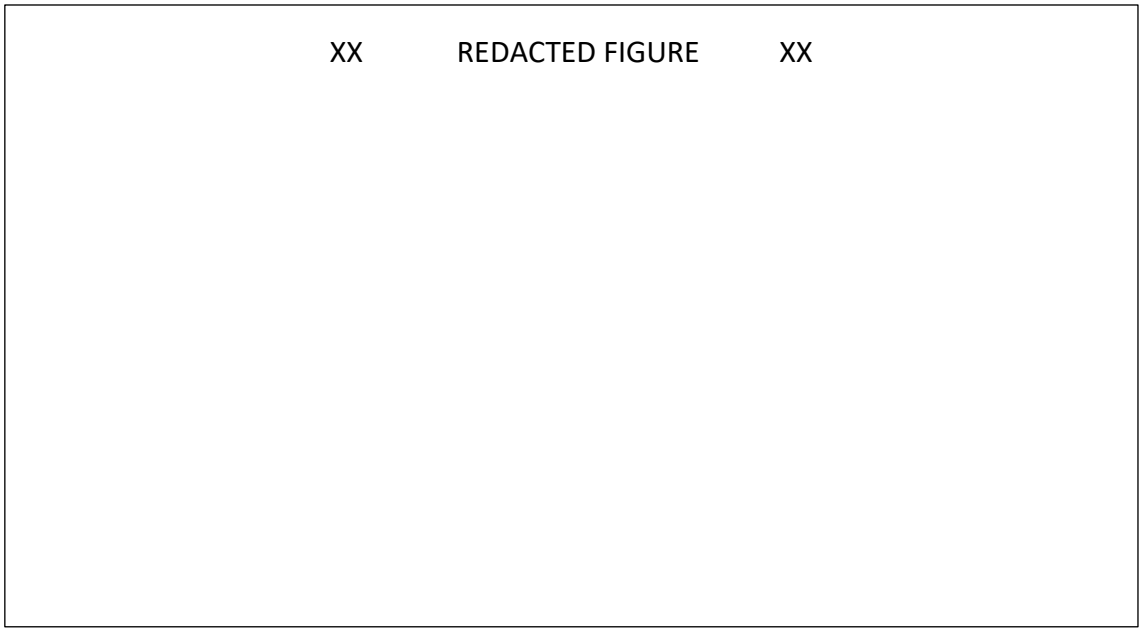

Figure 4.14. XX REDACTED TEXT XX.

XX REDACTED TEXT XX

XX REDACTED TEXT XX

XX REDACTED TABLE XX

Table 4.4. XX REDACTED TEXT XX.

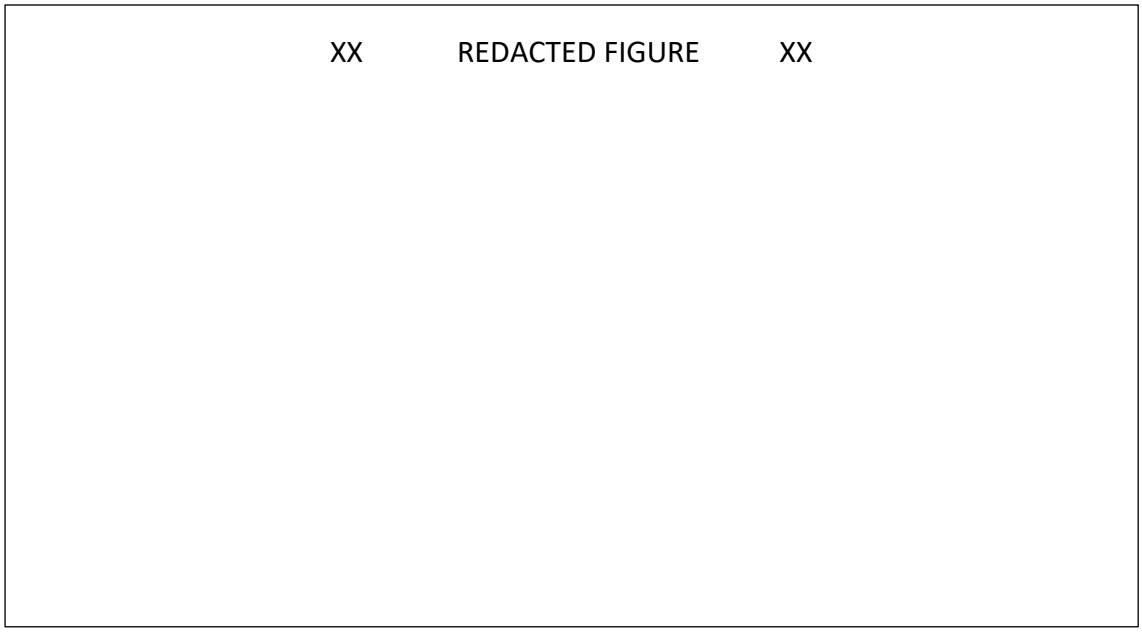

Figure 4.15. XX REDACTED TEXT XX.

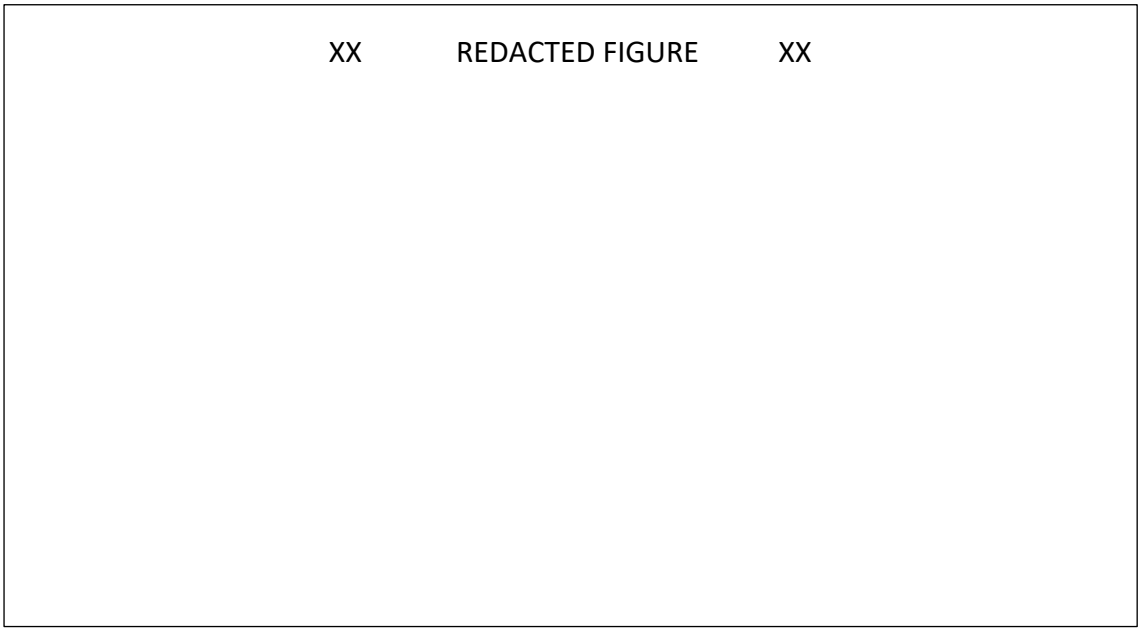

Figure 4.16. XX REDACTED TEXT XX.

XX REDACTED TEXT XX

XX REDACTED FIGURE XX

Figure 4.17. XX REDACTED TEXT XX.

XX REDACTED TEXT XX

XX REDACTED FIGURE XX

Figure 4.18. XX REDACTED TEXT XX.

XX REDACTED TABLE XX

XX REDACTED TABLE XX

CONTINUATION

Table 4.3. XX REDACTED TEXT XX.

XX REDACTED TABLE XX

XX REDACTED TABLE XX

CONTINUATION

XX REDACTED TABLE XX

CONTINUATION

Table 4.4. XX REDACTED TEXT XX.

XX REDACTED TABLE XX

XX REDACTED TABLE XX

CONTINUATION

XX REDACTED TABLE XX

CONTINUATION

Table 4.5. XX REDACTED TEXT XX.

XX REDACTED TEXT XX

XX REDACTED TABLE XX

Table 4.6. XX REDACTED TEXT XX.

4.3. XX REDACTED TEXT XX

4.3.1. Overview

XX REDACTED TEXT XX

4.3.2. Aims and objectives.

XX REDACTED TEXT XX

4.3.2.1. Objectives

- XX REDACTED TEXT XX

XX REDACTED TEXT XX

4.3.3. Results

XX REDACTED TEXT XX

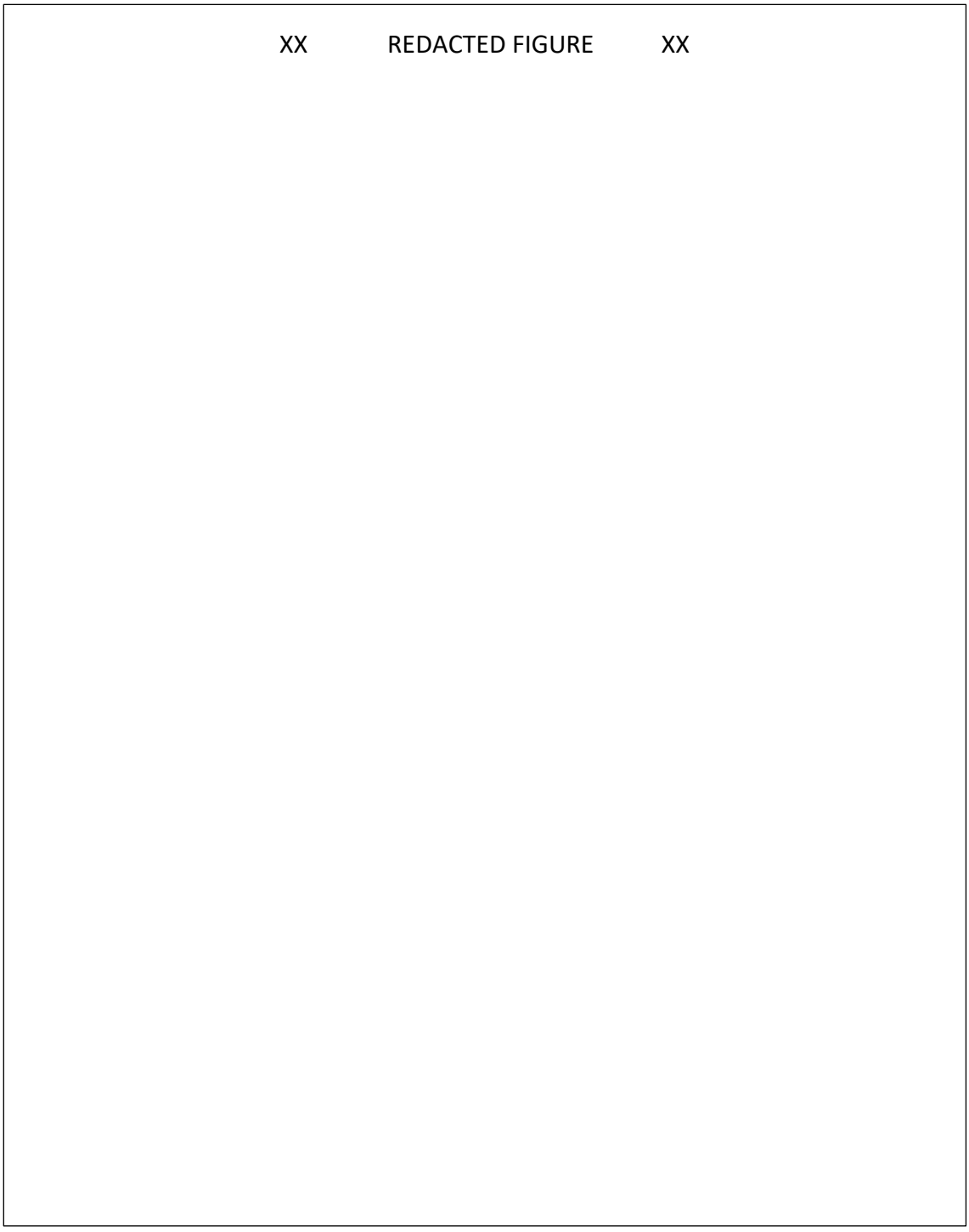

Figure 4.19. XX REDACTED TEXT XX.

XX REDACTED TEXT XX

XX REDACTED FIGURE XX

Figure 4.20. XX REDACTED TEXT XX.

XX REDACTED TEXT XX

XX REDACTED TEXT XX

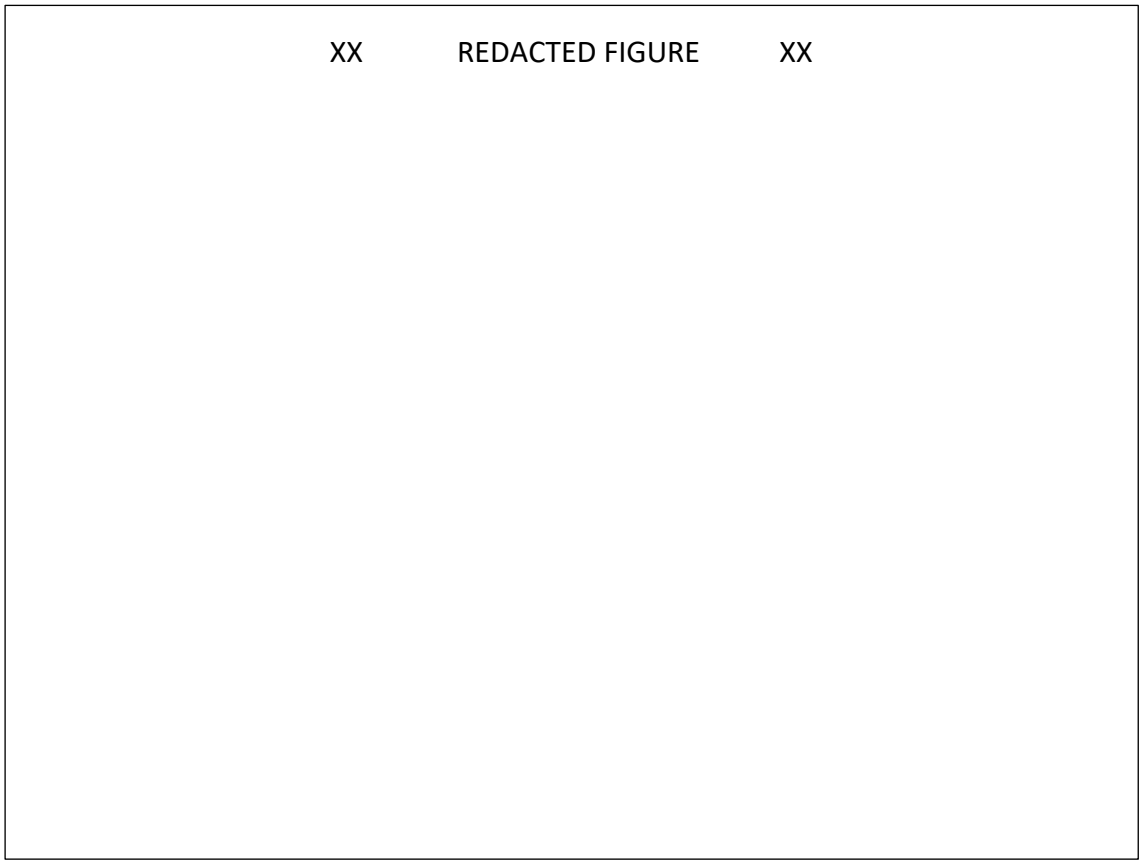

Figure 4.21. XX REDACTED TEXT XX.

4.4. DISCUSSION

4.4.1. XX REDACTED TEXT XX.

XX REDACTED TEXT XX

XX REDACTED TEXT XX

XX REDACTED TEXT XX

XX REDACTED TEXT XX

XX REDACTED TEXT XX

XX REDACTED TEXT XX

XX REDACTED TEXT XX

5. CHAPTER V: GENERAL DISCUSSION

5.1. Overview

Spinal muscular atrophies (SMAs) are described as hereditary neuromuscular disorders characterized by degeneration of lower motor neurons and are considered the main genetic cause of infant mortality (Ahmad et al., 2016; Farrar et al., 2015). The study of SMAs has represented a challenge since their first reports, which date to 1891 by Guido Werdnig who found degeneration of anterior horn cells during the autopsy of brothers that presented muscular weakness. Additional descriptions of the syndrome followed during the same decade by Jonah Hoffmann. Years later, Kugelberg, Welander and Dubowitz reported milder versions of the motor neuron condition described by Werdnig and Hoffman. Therefore, the classification of the different SMA types has been a subject of discussion over the years due to the variability between phenotypes.

The most common classification is based on age at onset symptoms (Mercuri, 2021). Further, over 30 genes with common pathophysiological pathways have been identified to be associated with these disorders; however, SMA usually refers to the most common form, which associates with insufficient levels of Survival Motor Neuron protein (SMN) expressed in motor neurons. Low SMN levels are caused by mutations or homozygous deletion of the Survival Motor Neuron 1 (*SMN1*) gene, which was identified about 100 years after the first SMA descriptions in the 1890's (Ahmad et al., 2016; Farrar et al., 2015; Lanfranco et al., 2017; Mercuri, 2021). Of note, *SMN2* a highly homologous gene also encodes for the SMN protein, albeit only approximately 10% of its transcripts generate a fully functional protein due to a mutation in exon 7 (Nicolau., et al 2021).

At a molecular level, the understanding of this disorder is even more challenging since SMN is a protein ubiquitously expressed, that has "housekeeping" and cell-type specific activities; nonetheless, the disease is characterised by neurodegeneration. This protein localizes in the cytoplasm and nucleus of all cells; and is highly present in brain, spinal cord, liver, and kidneys (Coady et al., 2011; Jangi et al., 2017; Li 2017; Renvoisé et al.,

2006). Further, SMN is suggested to be involved in a variety of cellular functions, including pre-mRNA splicing through the SMN complex; formation of small nuclear ribonucleoproteins (snRNPs) and small nucleolar RNPs; transcription during RNA pol II termination; in transport of mRNA within axons; the assembly of axonal granules; translation, among others. (Zhao et al., 2016; Jangi et al., 2017; Singh et al., 2018).

This plethora of SMN cellular roles inherently brings an elevated number of pathways and mechanisms that can be affected by SMN-deficiency. Previously, DNA damage has been observed in SMA in a general sense (Jangi et al., 2017; Kannan et al., 2018). Moreover, ribosomal DNA (rDNA) regions are prone to injury due to their high transcription rates; however, rDNA damage has not been a common line of study among SMA researchers (Moss et al., 2002). Therefore, in this thesis, I sought to explore R-loop-driven ribosomal DNA damage to determine its contribution to the physiopathology of the disease. Further, I searched for factors that interact with R-loops in an SMN-deficient context, to expand the scope of knowledge about SMA.

5.2. R-loops as a source of ribosomal DNA damage in SMA

Here, I studied a series of factors associated with DNA damage and genome instability in SMN-deficient cells. My results revealed elevated levels of γH2AX, showing that SMA patient cells accumulate DNA damage over ribosomal gene regions; suggested that rDNA damage could be a relevant contributor to the disease and that this occurs in all tissues, not only motor neurons.

XX REDACTED TEXT XX

To explain this, we know that the elevated levels of R-loops are associated with defects in R-loop resolution due to SMN deficiency as it has been demonstrated in previous papers (Crisitini et al., 2019) and in this thesis. However, the experiments carried out in

this project were limited to the comparison of two cell lines and this should be confirmed by testing more SMA patient fibroblast cell lines. Additionally, there is not much information in regards of why R-loops accumulate over rDNA as well. Meaning that it is not clear if SMN plays a role on a nucleolar level, or if its deficiency has an impact over other factors with nucleolar activity. Further, the mechanisms and proteins involved in rDNA R-loop resolution are less explored than those for nuclear R-loops.

XX REDACTED TEXT XX

The next item for discussion is the changes in the nucleoli on SMN-deficient cells, observed during the studies performed for this thesis. The nucleolar architecture is arranged according to ribosome biogenesis stages and depends on the status of rDNA transcription (van Sluis and McStay, 2017).

XX REDACTED TEXT XX

Accordingly, it is believed that rDNA molecules rearrange their configuration into structures like nucleolar caps, to facilitate repair. Moreover, protein factors could migrate in and out of the nucleoli. Therefore, these molecular rearrangements might have ultimately resulted in abnormal nucleoli in SMN-deficient cells. Importantly, these aberrant nucleoli were observed on both replicating and non-replicating cells. Such

reorganizations, possibly have a general impact in the nucleolar functions, like impaired rRNA synthesis and translation in neuronal cells as presented by Karyka et al. Interestingly, the decrease in rRNA synthesis and translation was only observed in SMN-deficient motor neurons and not cortical neurons (Karyka et al., 2022); thus, aligning with the tissue selectiveness of the disease.

5.3. DDX21 role as genome stability guardian in SMA

Among the proteins that normally reside in the nucleolar space, I studied the helicase DEAD box 21 (DDX21). According to NCBI database, DDX21 is mainly associated to watermelon stomach disease as an antigen recognized by autoimmune antibodies from a patient. Further, published literature have mentioned involvement of DDX21 as part of disorders such as some cancers and Treacher Collins syndrome. DDX21 interacts with the Pol I transcription complex in a helicase-dependent fashion and participates in R-loop resolution (Calo et al., 2015; Song et al., 2017). Additionally, DDX21 knockdown in epithelial cells revealed an increase in nucleolar R-loops, which contribute to rDNA damage (Koltowska et al., 2021). The Immunofluorescence analysis with α-DDX21 antibody revealed that in SMN-deficient cells DDX21 is displaced from the nucleoli to the nucleoplasm. Based on Calo et al. proposed mechanism, this migration could be triggered by the accumulation of rDNA damage signalling observed on ribosomal regions from SMA cells in this thesis. Therefore, in this scenario, if the presence of DDX21 in the nucleoli is associated with active transcription through its involvement with R-loop resolution, DDX21 nucleolar residence would be redundant if the transcription is stalled.

An intriguing relationship seems to exist between DDX21 and the tumour suppressor p53, where DDX21 has been reported to balance the function of p53. Loss of DDX21 in epithelial cells results in the activation of the p53 pathway (Koltowska et al., 2021). In Treacher Collins syndrome, mutations in elements from the transcriptional machinery of Pol I or its cofactor TCOF1 led to DDX21 displacement from the nucleoli to the nucleoplasm. However, blocking DDX21 migration rescued sensitivity to p53 and developmental defects (Calo et al., 2018).

In SMA patient cells, the accumulation of stable p53 occurs mainly in the nucleoli. Moreover, SMN and p53 interact *in vitro*, and they co-localise in cajal bodies upon p53 activation (Young et al., 2002). In SMA murine models, deficiency in SMN triggered activation of p53 in motor neurons. In particular, the phosphorylation of serine 18 in mouse was associated with motor neuron death (Simon et al., 2017). Nonetheless, several researchers agree that p53 is not the main driver for motor neuron death in mice since depletion of p53 did not avert motor neuron loss (Simon et al., 2017; Reedich et al., 2021; Buettner et al., 2022).

DDX21 could be having a double role in SMA by regulating R-loops levels and influencing cell death through its relationship with p53. It is interesting that both low SMN protein levels and loss of DDX21 result in R-loop accumulation and p53 activation. Also, that transcription disturbance by SMN-deficiency, mutations on RNA pol I or its cofactor TCOF1 cause migration of DDX21 form the nucleoli to the nucleoplasm. Nevertheless, is not clear the connection between DDX21 and SMN in normal conditions or their dynamic in SMA. DDX21 could be downstream from SMN activity and be downregulated or only displaced, resulting in R-loop accumulation; or a secondary mechanism such as DNA damage signalling (ATM, DNA-PKcs) could be causing DDX21 migration.

5.4. XX REDACTED TEXT XX.

XX REDACTED TEXT XX

XX REDACTED TEXT XX

XX REDACTED TEXT XX

5.5. Future directions

To continue the investigation of elements that contribute to the degeneration of motor neurons in Spinal Muscular Atrophy. First, I would probably carry further the study of rDNA damage by confirming my ChIP results in other SMA fibroblast cell lines. Also, see if it is possible to optimize an iPSC derived motor neuron ChIP or a similar experiment

to monitor ribosomal R-loop and γH2AX levels in neuronal cells and if feasible, compare those rates against the ones on dividing cells, to determine if these features are exacerbated on motor neurons.

XX REDACTED TEXT XX

Other interesting R-loop associated investigation might be the role of USP11 and SETX in the SMA context, since SETX is downregulated in SMA and depletion of USP11 and SETX leads to elevated nucleolar R-loops in MRC5 cells (Jurga et al., 2021).

XX REDACTED TEXT XX

The main purpose of the research carried out for this project as well as the future points of focus described in this section is to find out more about the molecular aspects of SMA. This way, the findings could potentially be translated into future therapies that could

supplement or substitute current treatments. So far nuclear R-loops have been considered interesting as one of the physiological alterations in SMA and if this thought is extended to nucleolar R-loops, it would open a possibility for the development of nucleolar R-loop regulators to ameliorate symptoms in SMA patients. In this sense DDX21 is one of the first candidates considered in this research as a nucleolar R-loop regulator, due to its association with active rDNA transcription and DNA/RNA helicase activity. It is possible that the role of DDX21 in the nucleoli might be similar to the one played by Senataxin (a stablished R-loop helicase) in the nucleus. The benefits of this type of treatments could potentially be the improvement of both neuronal and multi-organ symptoms since they would have a systemic impact on the human body. This is especially important nowadays that it has become more evident that not only the motor neurons are affected, but many other organs as well. Further, current treatments appear to help with the neuronal and muscular symptoms, but the rest of the organs keep deteriorating; thus, a systemic approach might have a significant positive impact on SMA patients' quality of life.

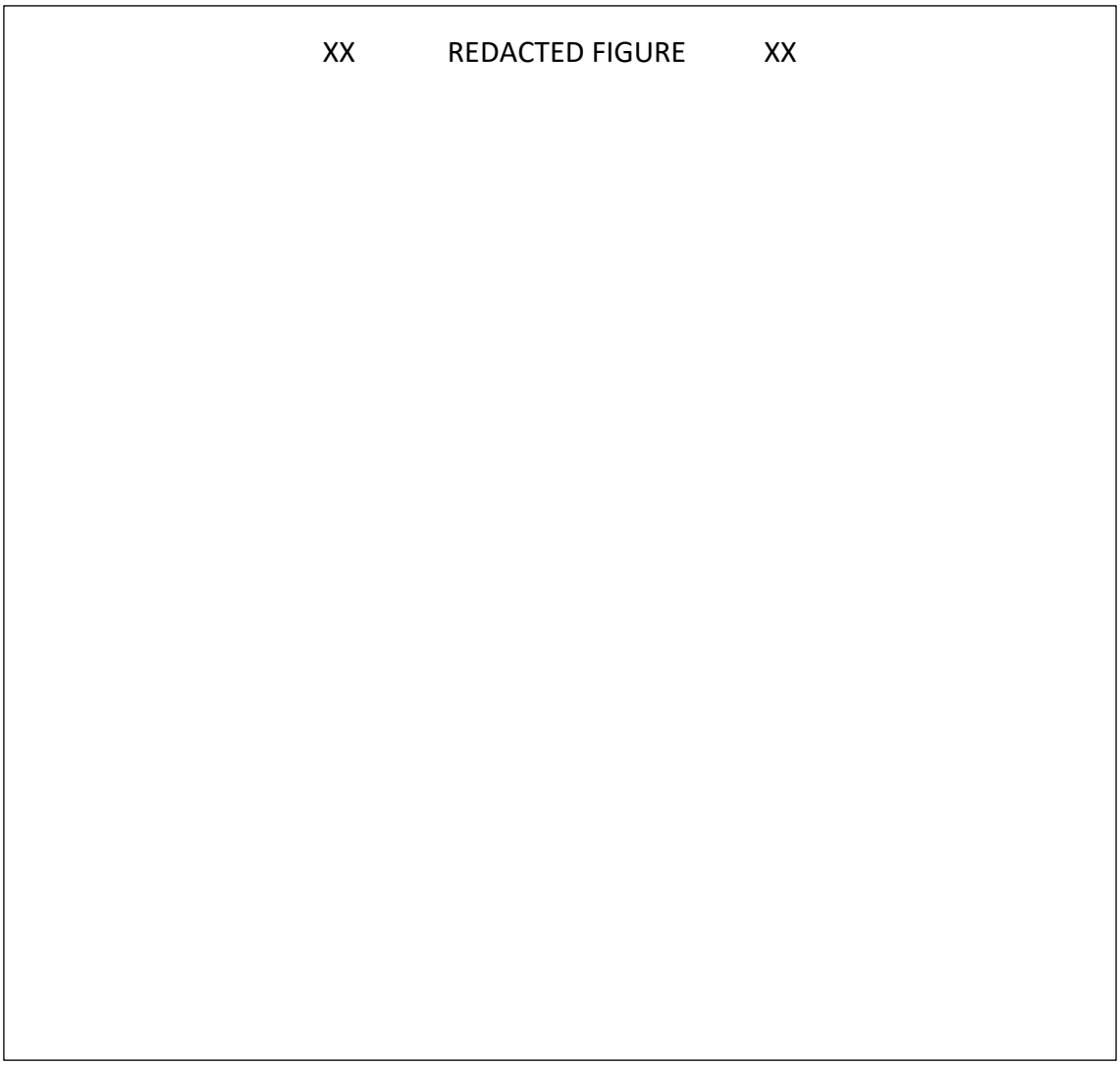

Figure 5.1 XX REDACTED TEXT XX.

6. REFERENCES

1. Abraham, K. J., Khosraviani, N., Chan, J. N. Y., Gorthi, A., Samman, A., Zhao, D. Y., Wang, M., Bokros, M., Vidya, E., Ostrowski, L. A., Oshidari, R., Pietrobon, V., Patel, P. S., Algouneh, A., Singhania, R., Liu, Y., Yerlici, V. T., De Carvalho, D. D., Ohh, M., Dickson, B. C., … Mekhail, K. (2020). Nucleolar RNA polymerase II drives ribosome biogenesis. *Nature*, *585*(7824), 298–302. https://doi.org/10.1038/s41586-020-2497-0
2. Adang L, Gavazzi F, De Simone M, Fazzi E, Galli J, Koh J, Kramer-Golinkoff J, De Giorgis V, Orcesi S, Peer K, Ulrick N, Woidill S, Shults J, Vanderver A. Developmental Outcomes of Aicardi Goutières Syndrome. J Child Neurol. 2020 Jan;35(1):7-16. doi: 10.1177/0883073819870944. Epub 2019 Sep 27. PMID: 31559893; PMCID: PMC7402202.
3. Aghamaleky Sarvestany, A., Hunter, G., Tavendale, A., Lamont, D. J., Llavero Hurtado, M., Graham, L. C., Wishart, T. M., & Gillingwater, T. H. (2014). Label-free quantitative proteomic profiling identifies disruption of ubiquitin homeostasis as a key driver of Schwann cell defects in spinal muscular atrophy. *Journal of proteome research*, *13*(11), 4546–4557. https://doi.org/10.1021/pr500492j
4. Ahmad, S., Bhatia, K., Kannan, A., & Gangwani, L. (2016). Molecular Mechanisms of Neurodegeneration in Spinal Muscular Atrophy. *Journal of experimental neuroscience*, *10*, 39–49. https://doi.org/10.4137/JEN.S33122
5. Alagoz, M., Chiang, S. C., Sharma, A., & El-Khamisy, S. F. (2013). ATM deficiency results in accumulation of DNA-topoisomerase I covalent intermediates in neural cells. *PloS one*, *8*(4), e58239. https://doi.org/10.1371/journal.pone.0058239
6. Andersen, J. S., Lam, Y. W., Leung, A. K., Ong, S. E., Lyon, C. E., Lamond, A. I., & Mann, M. (2005). Nucleolar proteome dynamics. *Nature*, *433*(7021), 77–83. https://doi.org/10.1038/nature03207
7. Arab, K., Karaulanov, E., Musheev, M., Trnka, P., Schäfer, A., Grummt, I., & Niehrs, C. (2019). GADD45A binds R-loops and recruits TET1 to CpG island promoters. *Nature genetics*, *51*(2), 217–223. https://doi.org/10.1038/s41588-018-0306-6
8. Backert S. (2002). R-loop-dependent rolling-circle replication and a new model for DNA concatemer resolution by mitochondrial plasmid mp1. *The EMBO journal*, *21*(12), 3128–3136. https://doi.org/10.1093/emboj/cdf311
9. Balk, B., Maicher, A., Dees, M., Klermund, J., Luke-Glaser, S., Bender, K., & Luke, B. (2013). Telomeric RNA-DNA hybrids affect telomere-length dynamics and senescence. *Nature structural & molecular biology*, *20*(10), 1199–1205. https://doi.org/10.1038/nsmb.2662
10. Beattie, C. E., & Kolb, S. J. (2018). Spinal muscular atrophy: Selective motor neuron loss and global defect in the assembly of ribonucleoproteins. *Brain research*, *1693*(Pt A), 92–97. https://doi.org/10.1016/j.brainres.2018.02.022
11. Beckett, D., Kovaleva, E., & Schatz, P. J. (1999). A minimal peptide substrate in biotin holoenzyme synthetase-catalyzed biotinylation. *Protein science : a publication of the Protein Society*, *8*(4), 921–929. https://doi.org/10.1110/ps.8.4.921
12. Bhatia, V., Barroso, S. I., García-Rubio, M. L., Tumini, E., Herrera-Moyano, E., & Aguilera, A. (2014). BRCA2 prevents R-loop accumulation and associates with TREX-

2 mRNA export factor PCID2. *Nature*, *511*(7509), 362–365. https://doi.org/10.1038/nature13374
13. Blokhina, Y. P., & Buchwalter, A. (2020). Moving fast and breaking things: Incidence and repair of DNA damage within ribosomal DNA repeats. *Mutation research*, *821*, 111715. https://doi.org/10.1016/j.mrfmmm.2020.111715
14. Boisvert, F. M., van Koningsbruggen, S., Navascués, J., & Lamond, A. I. (2007). The multifunctional nucleolus. *Nature reviews. Molecular cell biology*, *8*(7), 574–585. https://doi.org/10.1038/nrm2184
15. Boleslavska, B., Oravetzova, A., Shukla, K., Nascakova, Z., Ibini, O. N., Hasanova, Z., Andrs, M., Kanagaraj, R., Dobrovolna, J., & Janscak, P. (2022). DDX17 helicase promotes resolution of R-loop-mediated transcription-replication conflicts in human cells. *Nucleic acids research*, *50*(21), 12274–12290. https://doi.org/10.1093/nar/gkac1116
16. Bowerman, M., Murray, L. M., Scamps, F., Schneider, B. L., Kothary, R., & Raoul, C. (2018). Pathogenic commonalities between spinal muscular atrophy and amyotrophic lateral sclerosis: Converging roads to therapeutic development. *European journal of medical genetics*, *61*(11), 685–698. https://doi.org/10.1016/j.ejmg.2017.12.001
17. Boyd, P. J., Tu, W. Y., Shorrock, H. K., Groen, E. J. N., Carter, R. N., Powis, R. A., Thomson, S. R., Thomson, D., Graham, L. C., Motyl, A. A. L., Wishart, T. M., Highley, J. R., Morton, N. M., Becker, T., Becker, C. G., Heath, P. R., & Gillingwater, T. H. (2017). Bioenergetic status modulates motor neuron vulnerability and pathogenesis in a zebrafish model of spinal muscular atrophy. *PLoS genetics*, *13*(4), e1006744. https://doi.org/10.1371/journal.pgen.1006744
18. Buettner, J. M., Sowoidnich, L., Gerstner, F., Blanco-Redondo, B., Hallermann, S., & Simon, C. M. (2022). p53-dependent c-Fos expression is a marker but not executor for motor neuron death in spinal muscular atrophy mouse models. *Frontiers in cellular neuroscience*, *16*, 1038276. https://doi.org/10.3389/fncel.2022.1038276
19. Calo, E., Flynn, R. A., Martin, L., Spitale, R. C., Chang, H. Y., & Wysocka, J. (2015). RNA helicase DDX21 coordinates transcription and ribosomal RNA processing. *Nature*, *518*(7538), 249–253. https://doi.org/10.1038/nature13923
20. Calo, E., Gu, B., Bowen, M. E., Aryan, F., Zalc, A., Liang, J., Flynn, R. A., Swigut, T., Chang, H. Y., Attardi, L. D., & Wysocka, J. (2018). Tissue-selective effects of nucleolar stress and rDNA damage in developmental disorders. *Nature*, *554*(7690), 112–117. https://doi.org/10.1038/nature25449
21. Calucho, M., Bernal, S., Alías, L., March, F., Venceslá, A., Rodríguez-Álvarez, F. J., Aller, E., Fernández, R. M., Borrego, S., Millán, J. M., Hernández-Chico, C., Cuscó, I., Fuentes-Prior, P., & Tizzano, E. F. (2018). Correlation between SMA type and SMN2 copy number revisited: An analysis of 625 unrelated Spanish patients and a compilation of 2834 reported cases. *Neuromuscular disorders : NMD*, *28*(3), 208–215. https://doi.org/10.1016/j.nmd.2018.01.003
22. Calve, S., Witten, A., Ocken, A. et al. Incorporation of non-canonical amino acids into the developing murine proteome. *Sci Rep* 6, 32377 (2016). https://doi.org/10.1038/srep32377
23. Cargill, M., Venkataraman, R., & Lee, S. (2021). DEAD-Box RNA Helicases and Genome Stability. *Genes*, *12*(10), 1471. https://doi.org/10.3390/genes12101471

24. Cargill, M., Venkataraman, R., & Lee, S. (2021). DEAD-Box RNA Helicases and Genome Stability. *Genes*, *12*(10), 1471. https://doi.org/10.3390/genes12101471
25. Carusillo, A., & Mussolino, C. (2020). DNA Damage: From Threat to Treatment. *Cells*, *9*(7), 1665. https://doi.org/10.3390/cells9071665
26. Castillo-Guzman, D., & Chédin, F. (2021). Defining R-loop classes and their contributions to genome instability. *DNA repair*, *106*, 103182. https://doi.org/10.1016/j.dnarep.2021.103182
27. Ceccaldi, R., Sarangi, P., & D'Andrea, A. D. (2016). The Fanconi anaemia pathway: new players and new functions. *Nature reviews. Molecular cell biology*, *17*(6), 337–349. https://doi.org/10.1038/nrm.2016.48
28. Chakraborty, P., Huang, J. T. J., & Hiom, K. (2018). DHX9 helicase promotes R-loop formation in cells with impaired RNA splicing. *Nature communications*, *9*(1), 4346. https://doi.org/10.1038/s41467-018-06677-1
29. Chapman-Smith, A., & Cronan, J. E., Jr (1999). In vivo enzymatic protein biotinylation. *Biomolecular engineering*, *16*(1-4), 119–125. https://doi.org/10.1016/s1050-3862(99)00046-7
30. Chatterjee, N., & Walker, G. C. (2017). Mechanisms of DNA damage, repair, and mutagenesis. *Environmental and molecular mutagenesis*, *58*(5), 235–263. https://doi.org/10.1002/em.22087
31. Chaytow, H., Huang, Y. T., Gillingwater, T. H., & Faller, K. M. E. (2018). The role of survival motor neuron protein (SMN) in protein homeostasis. *Cellular and molecular life sciences: CMLS*, *75*(21), 3877–3894. https://doi.org/10.1007/s00018-018-2849-1
32. Chen, L., Chen, J. Y., Zhang, X., Gu, Y., Xiao, R., Shao, C., Tang, P., Qian, H., Luo, D., Li, H., Zhou, Y., Zhang, D. E., & Fu, X. D. (2017). R-ChIP Using Inactive RNase H Reveals Dynamic Coupling of R-loops with Transcriptional Pausing at Gene Promoters. *Molecular cell*, *68*(4), 745–757.e5. https://doi.org/10.1016/j.molcel.2017.10.008
33. Cho, K. F., Branon, T. C., Udeshi, N. D., Myers, S. A., Carr, S. A., & Ting, A. Y. (2020). Proximity labeling in mammalian cells with TurboID and split-TurboID. *Nature protocols*, *15*(12), 3971–3999. https://doi.org/10.1038/s41596-020-0399-0
34. Coady, T. H., & Lorson, C. L. (2011). SMN in spinal muscular atrophy and snRNP biogenesis. *Wiley interdisciplinary reviews. RNA*, *2*(4), 546–564. https://doi.org/10.1002/wrna.76
35. Cohen, S., Puget, N., Lin, Y. L., Clouaire, T., Aguirrebengoa, M., Rocher, V., Pasero, P., Canitrot, Y., & Legube, G. (2018). Senataxin resolves RNA:DNA hybrids forming at DNA double-strand breaks to prevent translocations. *Nature communications*, *9*(1), 533. https://doi.org/10.1038/s41467-018-02894-w
36. Crewe, M., & Madabhushi, R. (2021). Topoisomerase-Mediated DNA Damage in Neurological Disorders. *Frontiers in aging neuroscience*, *13*, 751742. https://doi.org/10.3389/fnagi.2021.751742
37. Cristini, A., Groh, M., Kristiansen, M. S., & Gromak, N. (2018). RNA/DNA Hybrid Interactome Identifies DXH9 as a Molecular Player in Transcriptional Termination and R-Loop-Associated DNA Damage. *Cell reports*, *23*(6), 1891–1905. https://doi.org/10.1016/j.celrep.2018.04.025
38. Cristini, A., Park, J. H., Capranico, G., Legube, G., Favre, G., & Sordet, O. (2016). DNA-PK triggers histone ubiquitination and signaling in response to DNA double-strand

breaks produced during the repair of transcription-blocking topoisomerase I lesions. *Nucleic acids research*, *44*(3), 1161–1178. https://doi.org/10.1093/nar/gkv1196
39. Cristini, A., Ricci, G., Britton, S., Salimbeni, S., Huang, S. N., Marinello, J., Calsou, P., Pommier, Y., Favre, G., Capranico, G., Gromak, N., & Sordet, O. (2019). Dual Processing of R-Loops and Topoisomerase I Induces Transcription-Dependent DNA Double-Strand Breaks. *Cell reports*, *28*(12), 3167–3181.e6. https://doi.org/10.1016/j.celrep.2019.08.041
40. Cristini, A., Tellier, M., Constantinescu, F., Accalai, C., Albulescu, L. O., Heiringhoff, R., Bery, N., Sordet, O., Murphy, S., & Gromak, N. (2022). RNase H2, mutated in Aicardi-Goutières syndrome, resolves co-transcriptional R-loops to prevent DNA breaks and inflammation. *Nature communications*, *13*(1), 2961. https://doi.org/10.1038/s41467-022-30604-0
41. Cuartas, J., & Gangwani, L. (2022). R-loop Mediated DNA Damage and Impaired DNA Repair in Spinal Muscular Atrophy. *Frontiers in cellular neuroscience*, *16*, 826608. https://doi.org/10.3389/fncel.2022.826608
42. Donlin-Asp, P. G., Bassell, G. J., & Rossoll, W. (2016). A role for the survival of motor neuron protein in mRNP assembly and transport. *Current opinion in neurobiology*, *39*, 53–61. https://doi.org/10.1016/j.conb.2016.04.004
43. Drees B. L. (1999). Progress and variations in two-hybrid and three-hybrid technologies. *Current opinion in chemical biology*, *3*(1), 64–70. https://doi.org/10.1016/s1367-5931(99)80012-x
44. Drolet, M., Bi, X., & Liu, L. F. (1994). Hypernegative supercoiling of the DNA template during transcription elongation in vitro. *The Journal of biological chemistry*, *269*(3), 2068–2074.
45. Du, Z. W., Chen, H., Liu, H., Lu, J., Qian, K., Huang, C. L., Zhong, X., Fan, F., & Zhang, S. C. (2015). Generation and expansion of highly pure motor neuron progenitors from human pluripotent stem cells. *Nature communications*, *6*, 6626. https://doi.org/10.1038/ncomms7626
46. Dull, T., Zufferey, R., Kelly, M., Mandel, R. J., Nguyen, M., Trono, D., & Naldini, L. (1998). A third-generation lentivirus vector with a conditional packaging system. *Journal of virology*, *72*(11), 8463–8471. https://doi.org/10.1128/JVI.72.11.8463-8471.1998
47. Dundas, C. M., Demonte, D., & Park, S. (2013). Streptavidin-biotin technology: improvements and innovations in chemical and biological applications. *Applied microbiology and biotechnology*, *97*(21), 9343–9353. https://doi.org/10.1007/s00253-013-5232-z
48. Escobar, J. S., Glémin, S., & Galtier, N. (2011). GC-biased gene conversion impacts ribosomal DNA evolution in vertebrates, angiosperms, and other eukaryotes. *Molecular biology and evolution*, *28*(9), 2561–2575. https://doi.org/10.1093/molbev/msr079
49. Farrar, M. A., & Kiernan, M. C. (2015). The Genetics of Spinal Muscular Atrophy: Progress and Challenges. *Neurotherapeutics : the journal of the American Society for Experimental NeuroTherapeutics*, *12*(2), 290–302. https://doi.org/10.1007/s13311-014-0314-x
50. Finkel, R. S., Mercuri, E., Meyer, O. H., Simonds, A. K., Schroth, M. K., Graham, R. J., Kirschner, J., Iannaccone, S. T., Crawford, T. O., Woods, S., Muntoni, F., Wirth, B., Montes, J., Main, M., Mazzone, E. S., Vitale, M., Snyder, B., Quijano-Roy, S., Bertini,

E., Davis, R. H., … SMA Care group (2018). Diagnosis and management of spinal muscular atrophy: Part 2: Pulmonary and acute care; medications, supplements and immunizations; other organ systems; and ethics. *Neuromuscular disorders : NMD*, *28*(3), 197–207. https://doi.org/10.1016/j.nmd.2017.11.004

51. García-Muse, T., & Aguilera, A. (2019). R Loops: From Physiological to Pathological Roles. *Cell,* 179(3), 604–618. https://doi.org/10.1016/j.cell.2019.08.055
52. García-Rubio, M. L., Pérez-Calero, C., Barroso, S. I., Tumini, E., Herrera-Moyano, E., Rosado, I. V., & Aguilera, A. (2015). The Fanconi Anemia Pathway Protects Genome Integrity from R-loops. *PLoS genetics*, *11*(11), e1005674. https://doi.org/10.1371/journal.pgen.1005674
53. Groen EJN, Gillingwater TH. UBA1: At the Crossroads of Ubiquitin Homeostasis and Neurodegeneration. Trends Mol Med. 2015 Oct;21(10):622-632. doi: 10.1016/j.molmed.2015.08.003. PMID: 26432019; PMCID: PMC4596250.
54. Groh, M., & Gromak, N. (2014). Out of balance: R-loops in human disease. *PLoS genetics*, *10*(9), e1004630. https://doi.org/10.1371/journal.pgen.1004630
55. Groh, M., Lufino, M. M., Wade-Martins, R., & Gromak, N. (2014). R-loops associated with triplet repeat expansions promote gene silencing in Friedreich ataxia and fragile X syndrome. *PLoS genetics*, *10*(5), e1004318. https://doi.org/10.1371/journal.pgen.1004318
56. Grunseich, C., Wang, I. X., Watts, J. A., Burdick, J. T., Guber, R. D., Zhu, Z., Bruzel, A., Lanman, T., Chen, K., Schindler, A. B., Edwards, N., Ray-Chaudhury, A., Yao, J., Lehky, T., Piszczek, G., Crain, B., Fischbeck, K. H., & Cheung, V. G. (2018). Senataxin Mutation Reveals How R-Loops Promote Transcription by Blocking DNA Methylation at Gene Promoters. *Molecular cell*, *69*(3), 426–437.e7. https://doi.org/10.1016/j.molcel.2017.12.030
57. Grzanka M, Piekiełko-Witkowska A. The Role of *TCOF1* Gene in Health and Disease: Beyond Treacher Collins Syndrome. Int J Mol Sci. 2021 Mar 1;22(5):2482. doi: 10.3390/ijms22052482. PMID: 33804586; PMCID: PMC7957619.
58. Gueiderikh, A., Maczkowiak-Chartois, F., & Rosselli, F. (2022). A new frontier in Fanconi anemia: From DNA repair to ribosome biogenesis. *Blood reviews*, *52*, 100904. https://doi.org/10.1016/j.blre.2021.100904
59. Halász, L., Karányi, Z., Boros-Oláh, B., Kuik-Rózsa, T., Sipos, É., Nagy, É., Mosolygó-L, Á., Mázló, A., Rajnavölgyi, É., Halmos, G., & Székvölgyi, L. (2017). RNA-DNA hybrid (R-loop) immunoprecipitation mapping: an analytical workflow to evaluate inherent biases. *Genome research*, *27*(6), 1063–1073. https://doi.org/10.1101/gr.219394.116
60. Harding, S. M., Boiarsky, J. A., & Greenberg, R. A. (2015). ATM Dependent Silencing Links Nucleolar Chromatin Reorganization to DNA Damage Recognition. *Cell reports*, *13*(2), 251–259. https://doi.org/10.1016/j.celrep.2015.08.085
61. Hatchi, E., Skourti-Stathaki, K., Ventz, S., Pinello, L., Yen, A., Kamieniarz-Gdula, K., Dimitrov, S., Pathania, S., McKinney, K. M., Eaton, M. L., Kellis, M., Hill, S. J., Parmigiani, G., Proudfoot, N. J., & Livingston, D. M. (2015). BRCA1 recruitment to transcriptional pause sites is required for R-loop-driven DNA damage repair. *Molecular cell*, *57*(4), 636–647. https://doi.org/10.1016/j.molcel.2015.01.011
62. Helena, J. M., Joubert, A. M., Grobbelaar, S., Nolte, E. M., Nel, M., Pepper, M. S., Coetzee, M., & Mercier, A. E. (2018). Deoxyribonucleic Acid Damage and Repair: Capitalizing on Our Understanding of the Mechanisms of Maintaining Genomic

Integrity for Therapeutic Purposes. *International journal of molecular sciences*, *19*(4), 1148. https://doi.org/10.3390/ijms19041148

63. Helmrich, A., Ballarino, M., & Tora, L. (2011). Collisions between replication and transcription complexes cause common fragile site instability at the longest human genes. *Molecular cell*, *44*(6), 966–977. https://doi.org/10.1016/j.molcel.2011.10.013
64. Hensel, N., & Claus, P. (2018). The Actin Cytoskeleton in SMA and ALS: How Does It Contribute to Motoneuron Degeneration?. *The Neuroscientist : a review journal bringing neurobiology, neurology and psychiatry*, *24*(1), 54–72. https://doi.org/10.1177/1073858417705059
65. Herold, S., Kalb, J., Büchel, G., Ade, C. P., Baluapuri, A., Xu, J., Koster, J., Solvie, D., Carstensen, A., Klotz, C., Rodewald, S., Schülein-Völk, C., Dobbelstein, M., Wolf, E., Molenaar, J., Versteeg, R., Walz, S., & Eilers, M. (2019). Recruitment of BRCA1 limits MYCN-driven accumulation of stalled RNA polymerase. *Nature*, *567*(7749), 545–549. https://doi.org/10.1038/s41586-019-1030-9
66. Hodroj, D., Serhal, K., & Maiorano, D. (2017). Ddx19 links mRNA nuclear export with progression of transcription and replication and suppresses genomic instability upon DNA damage in proliferating cells. *Nucleus (Austin, Tex.)*, *8*(5), 489–495. https://doi.org/10.1080/19491034.2017.1348448
67. Houston, K., Buschang, P. H., Iannaccone, S. T., & Seale, N. S. (1994). Craniofacial morphology of spinal muscular atrophy. *Pediatric research*, *36*(2), 265–269. https://doi.org/10.1203/00006450-199408000-00020
68. Huang, daW., Sherman, B. T., & Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature protocols*, *4*(1), 44–57. https://doi.org/10.1038/nprot.2008.211
69. Huang, S. N., Williams, J. S., Arana, M. E., Kunkel, T. A., & Pommier, Y. (2017). Topoisomerase I-mediated cleavage at unrepaired ribonucleotides generates DNA double-strand breaks. *The EMBO journal*, *36*(3), 361–373. https://doi.org/10.15252/embj.201592426
70. Huertas, P., & Aguilera, A. (2003). Cotranscriptionally formed DNA:RNA hybrids mediate transcription elongation impairment and transcription-associated recombination. *Molecular cell*, *12*(3), 711–721. https://doi.org/10.1016/j.molcel.2003.08.010
71. Hyjek, M., Figiel, M., & Nowotny, M. (2019). RNases H: Structure and mechanism. *DNA repair*, *84*, 102672. https://doi.org/10.1016/j.dnarep.2019.102672
72. Jangi, M., Fleet, C., Cullen, P., Gupta, S. V., Mekhoubad, S., Chiao, E., Allaire, N., Bennett, C. F., Rigo, F., Krainer, A. R., Hurt, J. A., Carulli, J. P., & Staropoli, J. F. (2017). SMN deficiency in severe models of spinal muscular atrophy causes widespread intron retention and DNA damage. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(12), E2347–E2356. https://doi.org/10.1073/pnas.1613181114
73. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science. 2012 Aug 17;337(6096):816-21. doi: 10.1126/science.1225829. Epub 2012 Jun 28. PMID: 22745249; PMCID: PMC6286148.
74. Ju, B. G., Lunyak, V. V., Perissi, V., Garcia-Bassets, I., Rose, D. W., Glass, C. K., & Rosenfeld, M. G. (2006). A topoisomerase IIbeta-mediated dsDNA break required for

regulated transcription. *Science (New York, N.Y.)*, *312*(5781), 1798–1802. https://doi.org/10.1126/science.1127196
75. Jurga, M., Abugable, A. A., Goldman, A. S. H., & El-Khamisy, S. F. (2021). USP11 controls R-loops by regulating senataxin proteostasis. *Nature communications*, *12*(1), 5156. https://doi.org/10.1038/s41467-021-25459-w
76. Kannan, A., Bhatia, K., Branzei, D., & Gangwani, L. (2018). Combined deficiency of Senataxin and DNA-PKcs causes DNA damage accumulation and neurodegeneration in spinal muscular atrophy. *Nucleic acids research*, *46*(16), 8326–8346. https://doi.org/10.1093/nar/gky641
77. Kannan, A., Cuartas, J., Gangwani, P., Branzei, D., & Gangwani, L. (2022). Mutation in senataxin alters the mechanism of R-loop resolution in amyotrophic lateral sclerosis 4. *Brain : a journal of neurology*, *145*(9), 3072–3094. https://doi.org/10.1093/brain/awab464
78. Kannan, A., Jiang, X., He, L., Ahmad, S., & Gangwani, L. (2020). ZPR1 prevents R-loop accumulation, upregulates SMN2 expression and rescues spinal muscular atrophy. *Brain : a journal of neurology*, *143*(1), 69–93. https://doi.org/10.1093/brain/awz373
79. Kantidze, O. L., Velichko, A. K., Luzhin, A. V., & Razin, S. V. (2016). Heat Stress-Induced DNA Damage. *Acta naturae*, *8*(2), 75–78.
80. Karahan, G., Sayar, N., Gozum, G., Bozkurt, B., Konu, O., & Yulug, I. G. (2015). Relative expression of rRNA transcripts and 45S rDNA promoter methylation status are dysregulated in tumors in comparison with matched-normal tissues in breast cancer. *Oncology reports*, *33*(6), 3131–3145. https://doi.org/10.3892/or.2015.3940
81. Karyka, E., Berrueta Ramirez, N., Webster, C. P., Marchi, P. M., Graves, E. J., Godena, V. K., Marrone, L., Bhargava, A., Ray, S., Ning, K., Crane, H., Hautbergue, G. M., El-Khamisy, S. F., & Azzouz, M. (2022). SMN-deficient cells exhibit increased ribosomal DNA damage. *Life science alliance*, *5*(8), e202101145. https://doi.org/10.26508/lsa.202101145
82. Katyal, S., Lee, Y., Nitiss, K. C., Downing, S. M., Li, Y., Shimada, M., Zhao, J., Russell, H. R., Petrini, J. H., Nitiss, J. L., & McKinnon, P. J. (2014). Aberrant topoisomerase-1 DNA lesions are pathogenic in neurodegenerative genome instability syndromes. *Nature neuroscience*, *17*(6), 813–821. https://doi.org/10.1038/nn.3715
83. Keinath MC, Prior DE, Prior TW. Spinal Muscular Atrophy: Mutations, Testing, and Clinical Relevance. Appl Clin Genet. 2021 Jan 25;14:11-25. doi: 10.2147/TACG.S239603. PMID: 33531827; PMCID: PMC7846873.
84. Killen, M. W., Stults, D. M., Adachi, N., Hanakahi, L., & Pierce, A. J. (2009). Loss of Bloom syndrome protein destabilizes human gene cluster architecture. *Human molecular genetics*, *18*(18), 3417–3428. https://doi.org/10.1093/hmg/ddp282
85. Kim, D. I., Birendra, K. C., Zhu, W., Motamedchaboki, K., Doye, V., & Roux, K. J. (2014). Probing nuclear pore complex architecture with proximity-dependent biotinylation. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(24), E2453–E2461. https://doi.org/10.1073/pnas.1406459111
86. Kim, D. S., Camacho, C. V., Nagari, A., Malladi, V. S., Challa, S., & Kraus, W. L. (2019). Activation of PARP-1 by snoRNAs Controls Ribosome Biogenesis and Cell Growth via the RNA Helicase DDX21. *Molecular cell*, *75*(6), 1270–1285.e14. https://doi.org/10.1016/j.molcel.2019.06.020
87. Kim, J. K., Jha, N. N., Feng, Z., Faleiro, M. R., Chiriboga, C. A., Wei-Lapierre, L., Dirksen, R. T., Ko, C. P., & Monani, U. R. (2020). Muscle-specific SMN reduction reveals motor

neuron-independent disease in spinal muscular atrophy models. *The Journal of clinical investigation*, *130*(3), 1271–1287. https://doi.org/10.1172/JCI131989
88. Kim, N., & Jinks-Robertson, S. (2012). Transcription as a source of genome instability. *Nature reviews. Genetics*, *13*(3), 204–214. https://doi.org/10.1038/nrg3152
89. Kiyomitsu, T., & Boerner, S. (2021). The Nuclear Mitotic Apparatus (NuMA) Protein: A Key Player for Nuclear Formation, Spindle Assembly, and Spindle Positioning. *Frontiers in cell and developmental biology*, *9*, 653801. https://doi.org/10.3389/fcell.2021.653801
90. Klusmann, I., Wohlberedt, K., Magerhans, A., Teloni, F., Korbel, J. O., Altmeyer, M., & Dobbelstein, M. (2018). Chromatin modifiers Mdm2 and RNF2 prevent RNA:DNA hybrids that impair DNA replication. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(48), E11311–E11320. https://doi.org/10.1073/pnas.1809592115
91. Koltowska, K., Okuda, K. S., Gloger, M., Rondon-Galeano, M., Mason, E., Xuan, J., Dudczig, S., Chen, H., Arnold, H., Skoczylas, R., Bower, N. I., Paterson, S., Lagendijk, A. K., Baillie, G. J., Leshchiner, I., Simons, C., Smith, K. A., Goessling, W., Heath, J. K., Pearson, R. B., … Hogan, B. M. (2021). The RNA helicase Ddx21 controls Vegfc-driven developmental lymphangiogenesis by balancing endothelial cell ribosome biogenesis and p53 function. *Nature cell biology*, *23*(11), 1136–1147. https://doi.org/10.1038/s41556-021-00784-w
92. Kong, L., Wang, X., Choe, D. W., Polley, M., Burnett, B. G., Bosch-Marcé, M., Griffin, J. W., Rich, M. M., & Sumner, C. J. (2009). Impaired synaptic vesicle release and immaturity of neuromuscular junctions in spinal muscular atrophy mice. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *29*(3), 842–851. https://doi.org/10.1523/JNEUROSCI.4434-08.2009
93. Kotsantis, P., Silva, L. M., Irmscher, S., Jones, R. M., Folkes, L., Gromak, N., & Petermann, E. (2016). Increased global transcription activity as a mechanism of replication stress in cancer. *Nature communications*, *7*, 13087. https://doi.org/10.1038/ncomms13087
94. Kwon, K., & Beckett, D. (2000). Function of a conserved sequence motif in biotin holoenzyme synthetases. *Protein science : a publication of the Protein Society*, *9*(8), 1530–1539. https://doi.org/10.1110/ps.9.8.1530
95. LaCava, J., Fernandez-Martinez, J., Hakhverdyan, Z., & Rout, M. P. (2016). Protein Complex Purification by Affinity Capture. *Cold Spring Harbor protocols*, *2016*(7), 10.1101/pdb.top077545. https://doi.org/10.1101/pdb.top077545
96. Lanfranco, M., Vassallo, N., & Cauchi, R. J. (2017). Spinal Muscular Atrophy: From Defective Chaperoning of snRNP Assembly to Neuromuscular Dysfunction. *Frontiers in molecular biosciences*, *4*, 41. https://doi.org/10.3389/fmolb.2017.00041
97. Lee, T. H., & Kang, T. H. (2019). DNA Oxidation and Excision Repair Pathways. *International journal of molecular sciences*, *20*(23), 6092. https://doi.org/10.3390/ijms20236092
98. Li, D. K., Tisdale, S., Lotti, F., & Pellizzoni, L. (2014). SMN control of RNP assembly: from post-transcriptional gene regulation to motor neuron disease. *Seminars in cell & developmental biology*, *32*, 22–29. https://doi.org/10.1016/j.semcdb.2014.04.026
99. Liao, C., Beveridge, R., Hudson, J. J. R., Parker, J. D., Chiang, S. C., Ray, S., Ashour, M. E., Sudbery, I., Dickman, M. J., & El-Khamisy, S. F. (2018). UCHL3 Regulates Topoisomerase-Induced Chromosomal Break Repair by Controlling TDP1

Proteostasis. *Cell reports*, *23*(11), 3352–3365. https://doi.org/10.1016/j.celrep.2018.05.033

100. Lin, C., Yang, L., Tanasa, B., Hutt, K., Ju, B. G., Ohgi, K., Zhang, J., Rose, D. W., Fu, X. D., Glass, C. K., & Rosenfeld, M. G. (2009). Nuclear receptor-induced chromosomal proximity and DNA breaks underlie specific translocations in cancer. *Cell*, *139*(6), 1069–1083. https://doi.org/10.1016/j.cell.2009.11.030
101. Lindström, M. S., Jurada, D., Bursac, S., Orsolic, I., Bartek, J., & Volarevic, S. (2018). Nucleolus as an emerging hub in maintenance of genome stability and cancer pathogenesis. *Oncogene*, *37*(18), 2351–2366. https://doi.org/10.1038/s41388-017-0121-z
102. Lipnick, S. L., Agniel, D. M., Aggarwal, R., Makhortova, N. R., Finlayson, S. G., Brocato, A., Palmer, N., Darras, B. T., Kohane, I., & Rubin, L. L. (2019). Systemic nature of spinal muscular atrophy revealed by studying insurance claims. *PloS one*, *14*(3), e0213680. https://doi.org/10.1371/journal.pone.0213680
103. Loomis, E. W., Sanz, L. A., Chédin, F., & Hagerman, P. J. (2014). Transcription-associated R-loop formation across the human FMR1 CGG-repeat region. *PLoS genetics*, *10*(4), e1004294. https://doi.org/10.1371/journal.pgen.1004294
104. Lu, B., Dong, L., Yi, D., Zhang, M., Zhu, C., Li, X., & Yi, C. (2020). Transposase-assisted tagmentation of RNA/DNA hybrid duplexes. *eLife*, *9*, e54919. https://doi.org/10.7554/eLife.54919
105. Majounie, E., Renton, A. E., Mok, K., Dopper, E. G., Waite, A., Rollinson, S., Chiò, A., Restagno, G., Nicolaou, N., Simon-Sanchez, J., van Swieten, J. C., Abramzon, Y., Johnson, J. O., Sendtner, M., Pamphlett, R., Orrell, R. W., Mead, S., Sidle, K. C., Houlden, H., Rohrer, J. D., … Traynor, B. J. (2012). Frequency of the C9orf72 hexanucleotide repeat expansion in patients with amyotrophic lateral sclerosis and frontotemporal dementia: a cross-sectional study. *The Lancet. Neurology*, *11*(4), 323–330. https://doi.org/10.1016/S1474-4422(12)70043-1
106. Manzo, S. G., Hartono, S. R., Sanz, L. A., Marinello, J., De Biasi, S., Cossarizza, A., Capranico, G., & Chedin, F. (2018). DNA Topoisomerase I differentially modulates R-loops across the human genome. *Genome biology*, *19*(1), 100. https://doi.org/10.1186/s13059-018-1478-1
107. Marinello, J., Bertoncini, S., Aloisi, I., Cristini, A., Malagoli Tagliazucchi, G., Forcato, M., Sordet, O., & Capranico, G. (2016). Dynamic Effects of Topoisomerase I Inhibition on R-Loops and Short Transcripts at Active Promoters. *PloS one*, *11*(1), e0147053. https://doi.org/10.1371/journal.pone.0147053
108. Marinello, J., Chillemi, G., Bueno, S., Manzo, S. G., & Capranico, G. (2013). Antisense transcripts enhanced by camptothecin at divergent CpG-island promoters associated with bursts of topoisomerase I-DNA cleavage complex and R-loop formation. *Nucleic acids research*, *41*(22), 10110–10123. https://doi.org/10.1093/nar/gkt778
109. May, W. A., Lessnick, S. L., Braun, B. S., Klemsz, M., Lewis, B. C., Lunsford, L. B., Hromas, R., & Denny, C. T. (1993). The Ewing's sarcoma EWS/FLI-1 fusion gene encodes a more potent transcriptional activator and is a more powerful transforming gene than FLI-1. *Molecular and cellular biology*, *13*(12), 7393–7398. https://doi.org/10.1128/mcb.13.12.7393-7398.1993
110. Mercuri E. (2021). Spinal muscular atrophy: from rags to riches. *Neuromuscular disorders : NMD*, *31*(10), 998–1003. https://doi.org/10.1016/j.nmd.2021.08.009

111. Mercuri, E., Finkel, R. S., Muntoni, F., Wirth, B., Montes, J., Main, M., Mazzone, E. S., Vitale, M., Snyder, B., Quijano-Roy, S., Bertini, E., Davis, R. H., Meyer, O. H., Simonds, A. K., Schroth, M. K., Graham, R. J., Kirschner, J., Iannaccone, S. T., Crawford, T. O., Woods, S., … SMA Care Group (2018). Diagnosis and management of spinal muscular atrophy: Part 1: Recommendations for diagnosis, rehabilitation, orthopedic and nutritional care. *Neuromuscular disorders : NMD*, *28*(2), 103–115. https://doi.org/10.1016/j.nmd.2017.11.005
112. Mersaoui, S. Y., Yu, Z., Coulombe, Y., Karam, M., Busatto, F. F., Masson, J. Y., & Richard, S. (2019). Arginine methylation of the DDX5 helicase RGG/RG motif by PRMT5 regulates resolution of RNA:DNA hybrids. *The EMBO journal*, *38*(15), e100986. https://doi.org/10.15252/embj.2018100986
113. Modesti, M. and Kanaar, R. (2001) DNA Repair: Spot(light)s on Chromatin. Current Biology, 11, R229-R232. http://dx.doi.org/10.1016/S0960-9822(01)00112-9
114. Moreira, M. C., & Koenig, M. (2004). Ataxia with Oculomotor Apraxia Type 2. In M. P. Adam (Eds.) et. al., *GeneReviews®*. University of Washington, Seattle.
115. Mosler, T., Conte, F., Longo, G. M. C., Mikicic, I., Kreim, N., Möckel, M. M., Petrosino, G., Flach, J., Barau, J., Luke, B., Roukos, V., & Beli, P. (2021). R-loop proximity proteomics identifies a role of DDX41 in transcription-associated genomic instability. *Nature communications*, *12*(1), 7314. https://doi.org/10.1038/s41467-021-27530-y
116. Moss, T., & Stefanovsky, V. Y. (2002). At the center of eukaryotic life. *Cell*, *109*(5), 545–548. https://doi.org/10.1016/s0092-8674(02)00761-4
117. Nguyen, H. H., Park, J., Kang, S., & Kim, M. (2015). Surface plasmon resonance: a versatile technique for biosensor applications. *Sensors (Basel, Switzerland)*, *15*(5), 10481–10510. https://doi.org/10.3390/s150510481
118. Nicolau, S., Waldrop, M. A., Connolly, A. M., & Mendell, J. R. (2021). Spinal Muscular Atrophy. *Seminars in pediatric neurology*, *37*, 100878. https://doi.org/10.1016/j.spen.2021.100878
119. Nölle, A., Zeug, A., van Bergeijk, J., Tönges, L., Gerhard, R., Brinkmann, H., Al Rayes, S., Hensel, N., Schill, Y., Apkhazava, D., Jablonka, S., O'mer, J., Srivastav, R. K., Baasner, A., Lingor, P., Wirth, B., Ponimaskin, E., Niedenthal, R., Grothe, C., & Claus, P. (2011). The spinal muscular atrophy disease protein SMN is linked to the Rho-kinase pathway via profilin. *Human molecular genetics*, *20*(24), 4865–4878. https://doi.org/10.1093/hmg/ddr425
120. Nowotny, M., Cerritelli, S. M., Ghirlando, R., Gaidamakov, S. A., Crouch, R. J., & Yang, W. (2008). Specific recognition of RNA/DNA hybrid and enhancement of human RNase H1 activity by HBD. *The EMBO journal*, 27(7), 1172–1181. https://doi.org/10.1038/emboj.2008.44
121. Okamoto, Y., Abe, M., Itaya, A., Tomida, J., Ishiai, M., Takaori-Kondo, A., Taoka, M., Isobe, T., & Takata, M. (2019). FANCD2 protects genome stability by recruiting RNA processing enzymes to resolve R-loops during mild replication stress. *The FEBS journal*, *286*(1), 139–150. https://doi.org/10.1111/febs.14700
122. Pan, L., Penney, J., & Tsai, L. H. (2014). Chromatin regulation of DNA damage repair and genome integrity in the central nervous system. *Journal of molecular biology*, *426*(20), 3376–3388. https://doi.org/10.1016/j.jmb.2014.08.001

123. Pećina-Šlaus, N., Kafka, A., Salamon, I., & Bukovac, A. (2020). Mismatch Repair Pathway, Genome Stability and Cancer. *Frontiers in molecular biosciences*, *7*, 122. https://doi.org/10.3389/fmolb.2020.00122
124. Pellizzoni, L., Charroux, B., Rappsilber, J., Mann, M., & Dreyfuss, G. (2001). A functional interaction between the survival motor neuron complex and RNA polymerase II. *The Journal of cell biology*, *152*(1), 75–85. https://doi.org/10.1083/jcb.152.1.75
125. Perego, M.G.L., Taiana, M., Bresolin, N. et al. R-Loops in Motor Neuron Diseases. *Mol Neurobiol* 56, 2579–2589 (2019). https://doi.org/10.1007/s12035-018-1246-y
126. Petermann, E., Lan, L., & Zou, L. (2022). Sources, resolution and physiological relevance of R-loops and RNA-DNA hybrids. *Nature reviews. Molecular cell biology*, *23*(8), 521–540. https://doi.org/10.1038/s41580-022-00474-x
127. Phillips, D. D., Garboczi, D. N., Singh, K., Hu, Z., Leppla, S. H., & Leysath, C. E. (2013). The sub-nanomolar binding of DNA-RNA hybrids by the single-chain Fv fragment of antibody S9.6. *Journal of molecular recognition : JMR*, *26*(8), 376–381. https://doi.org/10.1002/jmr.2284
128. Pommier, Y., Barcelo, J. M., Rao, V. A., Sordet, O., Jobson, A. G., Thibaut, L., Miao, Z. H., Seiler, J. A., Zhang, H., Marchand, C., Agama, K., Nitiss, J. L., & Redon, C. (2006). Repair of topoisomerase I-mediated DNA damage. *Progress in nucleic acid research and molecular biology*, *81*, 179–229. https://doi.org/10.1016/S0079-6603(06)81005-6
129. Pommier, Y., Redon, C., Rao, V. A., Seiler, J. A., Sordet, O., Takemura, H., Antony, S., Meng, L., Liao, Z., Kohlhagen, G., Zhang, H., & Kohn, K. W. (2003). Repair of and checkpoint response to topoisomerase I-mediated DNA damage. *Mutation research*, *532*(1-2), 173–203. https://doi.org/10.1016/j.mrfmmm.2003.08.016
130. Powell, W. T., Coulson, R. L., Gonzales, M. L., Crary, F. K., Wong, S. S., Adams, S., Ach, R. A., Tsang, P., Yamada, N. A., Yasui, D. H., Chédin, F., & LaSalle, J. M. (2013). R-loop formation at Snord116 mediates topotecan inhibition of Ube3a-antisense and allele-specific chromatin decondensation. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(34), 13938–13943. https://doi.org/10.1073/pnas.1305426110
131. Prior, T. W., Nagan, N., Sugarman, E. A., Batish, S. D., & Braastad, C. (2011). Technical standards and guidelines for spinal muscular atrophy testing. *Genetics in medicine: official journal of the American College of Medical Genetics*, *13*(7), 686–694. https://doi.org/10.1097/GIM.0b013e318220d523
132. Prozeller, D., Morsbach, S., & Landfester, K. (2019). Isothermal titration calorimetry as a complementary method for investigating nanoparticle-protein interactions. *Nanoscale*, *11*(41), 19265–19273. https://doi.org/10.1039/c9nr05790k
133. Raimer, A. C., Gray, K. M., & Matera, A. G. (2017). SMN - A chaperone for nuclear RNP social occasions?. *RNA biology*, *14*(6), 701–711. https://doi.org/10.1080/15476286.2016.1236168
134. Rastogi, R. P., Richa, Kumar, A., Tyagi, M. B., & Sinha, R. P. (2010). Molecular mechanisms of ultraviolet radiation-induced DNA damage and repair. *Journal of nucleic acids*, *2010*, 592980. https://doi.org/10.4061/2010/592980
135. Ray, S., Abugable, A. A., Parker, J., Liversidge, K., Palminha, N. M., Liao, C., Acosta-Martin, A. E., Souza, C. D. S., Jurga, M., Sudbery, I., & El-Khamisy, S. F. (2022). A

mechanism for oxidative damage repair at gene regulatory elements. *Nature*, *609*(7929), 1038–1047. https://doi.org/10.1038/s41586-022-05217-8

136. Ray, S., Panova, T., Miller, G., Volkov, A., Porter, A. C., Russell, J., Panov, K. I., & Zomerdijk, J. C. (2013). Topoisomerase IIα promotes activation of RNA polymerase I transcription by facilitating pre-initiation complex formation. *Nature communications*, 4, 1598. https://doi.org/10.1038/ncomms2599
137. Ray, S., Panova, T., Miller, G., Volkov, A., Porter, A. C., Russell, J., Panov, K. I., & Zomerdijk, J. C. (2013). Topoisomerase IIα promotes activation of RNA polymerase I transcription by facilitating pre-initiation complex formation. *Nature communications*, *4*, 1598. https://doi.org/10.1038/ncomms2599
138. Reedich, E. J., Kalski, M., Armijo, N., Cox, G. A., & DiDonato, C. J. (2021). Spinal motor neuron loss occurs through a p53-and-p21-independent mechanism in the $Smn^{2B/-}$ mouse model of spinal muscular atrophy. *Experimental neurology*, *337*, 113587. https://doi.org/10.1016/j.expneurol.2020.113587
139. Renvoisé, B., Khoobarry, K., Gendron, M. C., Cibert, C., Viollet, L., & Lefebvre, S. (2006). Distinct domains of the spinal muscular atrophy protein SMN are required for targeting to Cajal bodies in mammalian cells. *Journal of cell science*, *119*(Pt 4), 680–692. https://doi.org/10.1242/jcs.02782
140. Richard, P., & Manley, J. L. (2014). SETX sumoylation: A link between DNA damage and RNA surveillance disrupted in AOA2. *Rare diseases (Austin, Tex.)*, *2*, e27744. https://doi.org/10.4161/rdis.27744
141. Richard, P., & Manley, J. L. (2017). R Loops and Links to Human Disease. *Journal of molecular biology*, *429*(21), 3168–3180. https://doi.org/10.1016/j.jmb.2016.08.031
142. Rogakou, E. P., Boon, C., Redon, C., & Bonner, W. M. (1999). Megabase chromatin domains involved in DNA double-strand breaks in vivo. *The Journal of cell biology*, *146*(5), 905–916. https://doi.org/10.1083/jcb.146.5.905
143. Roux, K. J., Kim, D. I., Raida, M., & Burke, B. (2012). A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. *The Journal of cell biology*, *196*(6), 801–810. https://doi.org/10.1083/jcb.201112098
144. Salas-Armenteros I, Pérez-Calero C, Bayona-Feliu A, Tumini E, Luna R, Aguilera A. Human THO-Sin3A interaction reveals new mechanisms to prevent R-loops that cause genome instability. EMBO J. 2017 Dec 1;36(23):3532-3547. doi: 10.15252/embj.201797208. Epub 2017 Oct 26. PMID: 29074626; PMCID: PMC5709763.
145. Sanz, L. A., & Chédin, F. (2019). High-resolution, strand-specific R-loop mapping via S9.6-based DNA-RNA immunoprecipitation and high-throughput sequencing. *Nature protocols*, *14*(6), 1734–1755. https://doi.org/10.1038/s41596-019-0159-1
146. Sanz, L. A., Hartono, S. R., Lim, Y. W., Steyaert, S., Rajpurkar, A., Ginno, P. A., Xu, X., & Chédin, F. (2016). Prevalent, Dynamic, and Conserved R-Loop Structures Associate with Specific Epigenomic Signatures in Mammals. *Molecular cell*, *63*(1), 167–178. https://doi.org/10.1016/j.molcel.2016.05.032
147. Scheffer, H., Cobben, J. M., Matthijs, G., & Wirth, B. (2001). Best practice guidelines for molecular analysis in spinal muscular atrophy. *European journal of human genetics : EJHG*, *9*(7), 484–491. https://doi.org/10.1038/sj.ejhg.5200667
148. Schoefl, G. I. (1964). The effect of actinomycin D on the fine structure of thenucleolus.J. *Ultrastruct. Res*.10, 224–243. doi: 10.1016/s0022-5320(64)80007-1

149. Schorling, D. C., Pechmann, A., & Kirschner, J. (2020). Advances in Treatment of Spinal Muscular Atrophy - New Phenotypes, New Challenges, New Implications for Care. *Journal of neuromuscular diseases*, *7*(1), 1–13. https://doi.org/10.3233/JND-190424
150. Shababi, M., Lorson, C. L., & Rudnik-Schöneborn, S. S. (2014). Spinal muscular atrophy: a motor neuron disorder or a multi-organ disease?. *Journal of anatomy*, *224*(1), 15–28. https://doi.org/10.1111/joa.12083
151. Shen, W., Sun, H., De Hoyos, C. L., Bailey, J. K., Liang, X. H., & Crooke, S. T. (2017). Dynamic nucleoplasmic and nucleolar localization of mammalian RNase H1 in response to RNAP I transcriptional R-loops. *Nucleic acids research*, *45*(18), 10672–10692. https://doi.org/10.1093/nar/gkx710
152. Sherman, B. T., Hao, M., Qiu, J., Jiao, X., Baseler, M. W., Lane, H. C., Imamichi, T., & Chang, W. (2022). DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). *Nucleic acids research*, *50*(W1), W216–W221. Advance online publication. https://doi.org/10.1093/nar/gkac194
153. Simon, C. M., Dai, Y., Van Alstyne, M., Koutsioumpa, C., Pagiazitis, J. G., Chalif, J. I., Wang, X., Rabinowitz, J. E., Henderson, C. E., Pellizzoni, L., & Mentis, G. Z. (2017). Converging Mechanisms of p53 Activation Drive Motor Neuron Degeneration in Spinal Muscular Atrophy. *Cell reports*, *21*(13), 3767–3780. https://doi.org/10.1016/j.celrep.2017.12.003
154. Singh, R. N., Howell, M. D., Ottesen, E. W., & Singh, N. N. (2017). Diverse role of survival motor neuron protein. *Biochimica et biophysica acta. Gene regulatory mechanisms*, *1860*(3), 299–315. https://doi.org/10.1016/j.bbagrm.2016.12.008
155. Skourti-Stathaki, K., Kamieniarz-Gdula, K., & Proudfoot, N. J. (2014). R-loops induce repressive chromatin marks over mammalian gene terminators. *Nature*, *516*(7531), 436–439. https://doi.org/10.1038/nature13787
156. Skourti-Stathaki, K., Proudfoot, N. J., & Gromak, N. (2011). Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. *Molecular cell*, *42*(6), 794–805. https://doi.org/10.1016/j.molcel.2011.04.026
157. Sloan, K. E., Leisegang, M. S., Doebele, C., Ramírez, A. S., Simm, S., Safferthal, C., Kretschmer, J., Schorge, T., Markoutsa, S., Haag, S., Karas, M., Ebersberger, I., Schleiff, E., Watkins, N. J., & Bohnsack, M. T. (2015). The association of late-acting snoRNPs with human pre-ribosomal complexes requires the RNA helicase DDX21. *Nucleic acids research*, *43*(1), 553–564. https://doi.org/10.1093/nar/gku1291
158. Smolka, J. A., Sanz, L. A., Hartono, S. R., & Chédin, F. (2021). Recognition of RNA by the S9.6 antibody creates pervasive artifacts when imaging RNA:DNA hybrids. *The Journal of cell biology*, *220*(6), e202004079. https://doi.org/10.1083/jcb.202004079
159. Sollier, J., & Cimprich, K. A. (2015). Breaking bad: R-loops and genome integrity. *Trends in cell biology*, *25*(9), 514–522. https://doi.org/10.1016/j.tcb.2015.05.003
160. Sollier, J., Stork, C. T., García-Rubio, M. L., Paulsen, R. D., Aguilera, A., & Cimprich, K. A. (2014). Transcription-coupled nucleotide excision repair factors promote R-loop-induced genome instability. *Molecular cell*, *56*(6), 777–785. https://doi.org/10.1016/j.molcel.2014.10.020
161. Song, C., Hotz-Wagenblatt, A., Voit, R., & Grummt, I. (2017). SIRT7 and the DEAD-box helicase DDX21 cooperate to resolve genomic R loops and safeguard genome

stability. *Genes & development*, *31*(13), 1370–1381. https://doi.org/10.1101/gad.300624.117

162. Stabley, D. L., Harris, A. W., Holbrook, J., Chubbs, N. J., Lozo, K. W., Crawford, T. O., Swoboda, K. J., Funanage, V. L., Wang, W., Mackenzie, W., Scavina, M., Sol-Church, K., & Butchbach, M. E. (2015). SMN1 and SMN2 copy numbers in cell lines derived from patients with spinal muscular atrophy as measured by array digital PCR. *Molecular genetics & genomic medicine*, *3*(4), 248–257. https://doi.org/10.1002/mgg3.141
163. Stirling, P. C., & Hieter, P. (2017). Canonical DNA Repair Pathways Influence R-Loop-Driven Genome Instability. *Journal of molecular biology*, *429*(21), 3132–3138. https://doi.org/10.1016/j.jmb.2016.07.014
164. Sun, Q., Csorba, T., Skourti-Stathaki, K., Proudfoot, N. J., & Dean, C. (2013). R-loop stabilization represses antisense transcription at the Arabidopsis FLC locus. *Science (New York, N.Y.)*, *340*(6132), 619–621. https://doi.org/10.1126/science.1234848
165. Sun, Y., McCorvie, T. J., Yates, L. A., & Zhang, X. (2020). Structural basis of homologous recombination. *Cellular and molecular life sciences : CMLS*, *77*(1), 3–18. https://doi.org/10.1007/s00018-019-03365-1
166. Szklarczyk, D., Kirsch, R., Koutrouli, M., Nastou, K., Mehryary, F., Hachilif, R., Gable, A. L., Fang, T., Doncheva, N. T., Pyysalo, S., Bork, P., Jensen, L. J., & von Mering, C. (2023). The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic acids research*, *51*(D1), D638–D646. https://doi.org/10.1093/nar/gkac1000
167. Tan-Wong, S. M., Dhir, S., & Proudfoot, N. J. (2019). R-Loops Promote Antisense Transcription across the Mammalian Genome. *Molecular cell*, *76*(4), 600–616.e6. https://doi.org/10.1016/j.molcel.2019.10.002
168. Tan, J., Wang, X., Phoon, L., Yang, H., & Lan, L. (2020). Resolution of ROS-induced G-quadruplexes and R-loops at transcriptionally active sites is dependent on BLM helicase. *FEBS letters*, *594*(9), 1359–1367. https://doi.org/10.1002/1873-3468.13738
169. Tanaka, A., Wang, J. Y., Shia, J., Zhou, Y., Ogawa, M., Hendrickson, R. C., Klimstra, D. S., & Roehrl, M. H. (2020). DEAD-box RNA helicase protein DDX21 as a prognosis marker for early stage colorectal cancer with microsatellite instability. *Scientific reports*, *10*(1), 22085. https://doi.org/10.1038/s41598-020-79049-9
170. Tedeschi, F. A., Cloutier, S. C., Tran, E. J., & Jankowsky, E. (2018). The DEAD-box protein Dbp2p is linked to noncoding RNAs, the helicase Sen1p, and R-loops. *RNA (New York, N.Y.)*, *24*(12), 1693–1705. https://doi.org/10.1261/rna.067249.118
171. Teoh, H. L., Carey, K., Sampaio, H., Mowat, D., Roscioli, T., & Farrar, M. (2017). Inherited Paediatric Motor Neuron Disorders: Beyond Spinal Muscular Atrophy. *Neural plasticity*, *2017*, 6509493. https://doi.org/10.1155/2017/6509493
172. Tews, D. S., and Goebel, H. H. (1996). DNA fragmentation and BCL-2 expression in infantile spinal muscular atrophy. *Neuromuscul. Disord*.6, 265–273.doi: 10.1016/0960-8966(96)00018-1
173. Thomas, M., White, R. L., & Davis, R. W. (1976). Hybridization of RNA to double-stranded DNA: formation of R-loops. *Proceedings of the National Academy of Sciences of the United States of America*, *73*(7), 2294–2298. https://doi.org/10.1073/pnas.73.7.2294

174. Toubiana, S., & Selig, S. (2018). DNA:RNA hybrids at telomeres - when it is better to be out of the (R) loop. *The FEBS journal*, *285*(14), 2552–2566. https://doi.org/10.1111/febs.14464
175. Tresini, M., Warmerdam, D. O., Kolovos, P., Snijder, L., Vrouwe, M. G., Demmers, J. A., van IJcken, W. F., Grosveld, F. G., Medema, R. H., Hoeijmakers, J. H., Mullenders, L. H., Vermeulen, W., & Marteijn, J. A. (2015). The core spliceosome as target and effector of non-canonical ATM signalling. *Nature*, *523*(7558), 53–58. https://doi.org/10.1038/nature14512
176. Trinkle-Mulcahy L. (2019). Recent advances in proximity-based labeling methods for interactome mapping. *F1000Research*, *8*, F1000 Faculty Rev-135. https://doi.org/10.12688/f1000research.16903.1
177. Tubbs, A., & Nussenzweig, A. (2017). Endogenous DNA Damage as a Source of Genomic Instability in Cancer. *Cell*, *168*(4), 644–656. https://doi.org/10.1016/j.cell.2017.01.002
178. van Blitterswijk, M., Mullen, B., Heckman, M. G., Baker, M. C., DeJesus-Hernandez, M., Brown, P. H., Murray, M. E., Hsiung, G. Y., Stewart, H., Karydas, A. M., Finger, E., Kertesz, A., Bigio, E. H., Weintraub, S., Mesulam, M., Hatanpaa, K. J., White, C. L., 3rd, Neumann, M., Strong, M. J., Beach, T. G., … Rademakers, R. (2014). Ataxin-2 as potential disease modifier in C9ORF72 expansion carriers. *Neurobiology of aging*, *35*(10), 2421.e13–2421.e2.421E17. https://doi.org/10.1016/j.neurobiolaging.2014.04.016
179. van Sluis, M., & McStay, B. (2017). Nucleolar reorganization in response to rDNA damage. *Current opinion in cell biology*, *46*, 81–86. https://doi.org/10.1016/j.ceb.2017.03.004
180. van Steensel, B., & Henikoff, S. (2000). Identification of in vivo DNA targets of chromatin proteins using tethered dam methyltransferase. *Nature biotechnology*, *18*(4), 424–428. https://doi.org/10.1038/74487
181. Vohhodina J, Goehring LJ, Liu B, Kong Q, Botchkarev VV Jr, Huynh M, Liu Z, Abderazzaq FO, Clark AP, Ficarro SB, Marto JA, Hatchi E, Livingston DM. BRCA1 binds TERRA RNA and suppresses R-Loop-based telomeric DNA damage. Nat Commun. 2021 Jun 10;12(1):3542. doi: 10.1038/s41467-021-23716-6. PMID: 34112789; PMCID: PMC8192922.
182. Walker, C., Herranz-Martin, S., Karyka, E., Liao, C., Lewis, K., Elsayed, W., Lukashchuk, V., Chiang, S. C., Ray, S., Mulcahy, P. J., Jurga, M., Tsagakis, I., Iannitti, T., Chandran, J., Coldicott, I., De Vos, K. J., Hassan, M. K., Higginbottom, A., Shaw, P. J., Hautbergue, G. M., … El-Khamisy, S. F. (2017). C9orf72 expansion disrupts ATM-mediated chromosomal break repair. *Nature neuroscience*, *20*(9), 1225–1235. https://doi.org/10.1038/nn.4604
183. Wang, K., Wang, H., Li, C., Yin, Z., Xiao, R., Li, Q., Xiang, Y., Wang, W., Huang, J., Chen, L., Fang, P., & Liang, K. (2021). Genomic profiling of native R loops with a DNA-RNA hybrid recognition sensor. *Science advances*, *7*(8), eabe3516. https://doi.org/10.1126/sciadv.abe3516
184. Warmerdam, D. O., & Wolthuis, R. M. F. (2019). Keeping ribosomal DNA intact: a repeating challenge. *Chromosome research : an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology*, *27*(1-2), 57–72. https://doi.org/10.1007/s10577-018-9594-z

185. Wirth B. (2021). Spinal Muscular Atrophy: In the Challenge Lies a Solution. *Trends in neurosciences*, *44*(4), 306–322. https://doi.org/10.1016/j.tins.2020.11.009
186. Wishart TM, Mutsaers CA, Riessland M, Reimer MM, Hunter G, Hannam ML, Eaton SL, Fuller HR, Roche SL, Somers E, Morse R, Young PJ, Lamont DJ, Hammerschmidt M, Joshi A, Hohenstein P, Morris GE, Parson SH, Skehel PA, Becker T, Robinson IM, Becker CG, Wirth B, Gillingwater TH. Dysregulation of ubiquitin homeostasis and β-catenin signaling promote spinal muscular atrophy. J Clin Invest. 2014 Apr;124(4):1821-34. doi: 10.1172/JCI71318. Epub 2014 Mar 3. PMID: 24590288; PMCID: PMC3973095.
187. Wood, M., Quinet, A., Lin, Y. L., Davis, A. A., Pasero, P., Ayala, Y. M., & Vindigni, A. (2020). TDP-43 dysfunction results in R-loop accumulation and DNA replication defects. *Journal of cell science*, *133*(20), jcs244129. https://doi.org/10.1242/jcs.244129
188. Wu H, Sun H, Liang X, Lima WF, Crooke ST. Human RNase H1 is associated with protein P32 and is involved in mitochondrial pre-rRNA processing. PLoS One. 2013 Aug 22;8(8):e71006. doi: 10.1371/journal.pone.0071006. PMID: 23990920; PMCID: PMC3750045.
189. Wulfridge, P., & Sarma, K. (2021). A nuclease- and bisulfite-based strategy captures strand-specific R-loops genome-wide. *eLife*, *10*, e65146. https://doi.org/10.7554/eLife.65146
190. Xing, Y. H., Yao, R. W., Zhang, Y., Guo, C. J., Jiang, S., Xu, G., Dong, R., Yang, L., & Chen, L. L. (2017). SLERT Regulates DDX21 Rings Associated with Pol I Transcription. *Cell*, *169*(4), 664–678.e16. https://doi.org/10.1016/j.cell.2017.04.011
191. Yan, Q., Shields, E. J., Bonasio, R., & Sarma, K. (2019). Mapping Native R-Loops Genome-wide Using a Targeted Nuclease Approach. *Cell reports*, *29*(5), 1369–1380.e5. https://doi.org/10.1016/j.celrep.2019.09.052
192. Young, P. J., Day, P. M., Zhou, J., Androphy, E. J., Morris, G. E., & Lorson, C. L. (2002). A direct interaction between the survival motor neuron protein and p53 and its relationship to spinal muscular atrophy. *The Journal of biological chemistry*, *277*(4), 2852–2859. https://doi.org/10.1074/jbc.M108769200
193. Yousefzadeh, M., Henpita, C., Vyas, R., Soto-Palma, C., Robbins, P., & Niedernhofer, L. (2021). DNA damage-how and why we age?. *eLife*, *10*, e62852. https://doi.org/10.7554/eLife.62852
194. Yu, K., Chedin, F., Hsieh, C. L., Wilson, T. E., & Lieber, M. R. (2003). R-loops at immunoglobulin class switch regions in the chromosomes of stimulated B cells. *Nature immunology*, *4*(5), 442–451. https://doi.org/10.1038/ni919
195. Zhang, Y., Baysac, K. C., Yee, L. F., Saporita, A. J., & Weber, J. D. (2014). Elevated DDX21 regulates c-Jun activity and rRNA processing in human breast cancers. *Breast cancer research : BCR*, *16*(5), 449. https://doi.org/10.1186/s13058-014-0449-z
196. Zhang, Z., Lotti, F., Dittmar, K., Younis, I., Wan, L., Kasim, M., & Dreyfuss, G. (2008). SMN deficiency causes tissue-specific perturbations in the repertoire of snRNAs and widespread defects in splicing. *Cell*, *133*(4), 585–600. https://doi.org/10.1016/j.cell.2008.03.031
197. Zhao, B., Rothenberg, E., Ramsden, D. A., & Lieber, M. R. (2020). The molecular basis and disease relevance of non-homologous DNA end joining. *Nature reviews. Molecular cell biology*, *21*(12), 765–781. https://doi.org/10.1038/s41580-020-00297-8

198. Zhao, D. Y., Gish, G., Braunschweig, U., Li, Y., Ni, Z., Schmitges, F. W., Zhong, G., Liu, K., Li, W., Moffat, J., Vedadi, M., Min, J., Pawson, T. J., Blencowe, B. J., & Greenblatt, J. F. (2016). SMN and symmetric arginine dimethylation of RNA polymerase II C-terminal domain control termination. *Nature*, *529*(7584), 48–53. https://doi.org/10.1038/nature16469

199. Zhou, H., Li, L., Wang, Q., Hu, Y., Zhao, W., Gautam, M., & Li, L. (2020). H3K9 Demethylation-Induced R-Loop Accumulation Is Linked to Disorganized Nucleoli. *Frontiers in genetics*, *11*, 43. https://doi.org/10.3389/fgene.2020.00043

7. ANNEX

siRNA	Target Sequences (5'->3')
ON-TARGETplus Human SMN1 siRNA SMARTpool, Dharmacon™	GCUCACAUUCCUUAAAUUA GCUGAUGCUUUGGGAAGUA GAUGAUACAGCACUGAUAA UAAGCAUGCUCUAAAGAAU
Scrambled siRNA, eurofins	AGGUAGUGUAAUCGCCUUGUU
ON-TARGETplus Human DDX21 (9188) siRNA SMARTpool, Dharmacon™	Accession hits: NM_001256910, XM_011540336, NM_004728, XM_017016910

Table 7.1. List of siRNAs and their target sequences.

Primers used for qPCR	Sequence
Actin 5'pause (F)	TTA CCC AGA GTG CAG GTG TG
Actin 5'pause (R)	CCC CAA TAA GCA GGA ACA GA
PRR5L (F)	ACA TCA CCC ATG ATT ATC CCG TTA
PRR5L (R)	GAC ACA ACA GGC TTC CAG GTC
ING3 (F)	TTT TTC TTC TCT AAC TAC CCT CCC C
ING3 (R)	GTG CCC TAA TCT GAA TGA CTA CA
c-FOS3 (F)	TGA GCC CGT GAC GTT TAC
c-FOS3 (R)	TGC AGA TGC GGT TGG AG
SPRY2 in1A (F)	CGA ATT CGG CGC TGA GAG
SPRY2 in1A (R)	ACA GGT TAG AAA TGC GGG
R#3 (F)	AAA AGC CTT CTC TAG CGA TCT G
R#3 (R)	CAT AAC GGA GGC AGA GAC AG

R#5 (F)	GCG AGA GCC GAG AAC TCG
R#5 (R)	ACC AAC GAC ACG CCC TTC
R#7 (F)	GAC ACT TCG AAC GCA CTT G
R#7 (R)	CTC AGA CAG GCG TAG CCC CG
RPL32 (F)	GAAGTTCCTGGTCCACAACG
RPL32 (R)	GCGATCTCGGCACAGTAAG
18 S (F)	ATGGCCGTTCTTAGTTGGTG
18S (R)	CGCTGAGCCAGTCAGTGTAG
5.8S (F)	GACTCTTAGCGGTGGATCACTC
5.8S (R)	GACGCTCAGACAGGCGTAG
28S (F)	CAGGGGAATCCGACTGTTTA
28S (R)	ATGACGAGGCATTTGGCTAC
RPL13A (F)	AGGTGCCTTGCTCACAGAGT
RPL13A (R)	GGTTGCATTGCCCTCATTAC
CALM3 (F)	GAGGAATTGTGGCGTTGACT
CALM3 (R)	AGAGTGGCCAAATGAGCAGT
SNRPN (F) (negative)	GCCAAATGAGTGAGGATGGT
SNRPN (R) (negative)	TCCTCTCTGCCTGACTCCAT
EGR1 (F) (negative)	GAACGTTCAGCCTCGTTCTC
EGR1 (R) (negative)	GGAAGGTGGAAGGAAACACA
TFPT (F)	TCTGGGAGTCCAAGCAGACT
TFPT (R)	AAGGAGCCACTGAAGGGTTT

Table 7.2. List of primers used for qPCR and their nucleotide sequences.

Primers	Sequence
HB-GFP2 (F)	CGAGAATTCACG CGTCGCCGATGT TCTATGCCGTGAGG
HB-GFP2 (R)	CCGGGCTGCAGC ATATGCGCTTAC TTGTACAGCTCGTC
Lenti2 (F)	GCGCATATG CTGCAGCCC
Lenti2 (R)	CGGCGACGC GTGAATTCTC
LV sequencing (F)	GACCTCTAGGATCCTAATAC
LV sequencing (R)	GATTATCGGAATTCCCTCGA
HB domain (F)	cgagctcggtaccaagcttaATGTTCTATGCCGTGAGG
HB domain (R)	ggcacagtcgaggctgatcaTCAGCTTGCAGATTTCCTG

Table 7.3. List of primers used for plasmid generation and their nucleotide sequences.

Plasmids	Remarks
pMD.2G	Plasmid encoding the vesicular stomatitis virus G envelope
pCMVΔR8.92	Packaging plasmid encoding all the viral genes needed in trans.
pRSV-Rev	Plasmid encoding the rev protein of HIV-1.
pLenti-VOS	Lentiviral vector with Multiple Cloning Site (MCS)
pLVHB-EGFP	Transfer plasmid for the cloning of HB-EGPF. From this thesis.
pEGFP-N1-HB	From Sherif El-Khamisy.
pEGFP-N1	Clontech
REDACTED	REDACTED
REDACTED	REDACTED
pLVTHMshSMN1	(Human) From Mimoun Azzouz Oligo sequences

	Forward CGCGT CCCC (GGACAAATGTTCTGCCATT) TTCAAGAGA (AATGGCAGAACATTTGTCC) TTTTT GGAA AT Reverse CGAT TTCC AAAAA (GGACAAATGTTCTGCCATT) TCTCTTGAA (AATGGCAGAACATTTGTCC) GGGG A
pLVTHMshSMN2	(Human)From Mimoun Azzouz <u>Oligo sequences</u> Forward CGCGT CCCC (CAATCTGTGAAGTAGCTAA) TTCAAGAGA (TTAGCTACTTCACAGATTG) TTTTT GGAA AT Reverse CGAT TTCC AAAAA (CAATCTGTGAAGTAGCTAA) TCTCTTGAA (TTAGCTACTTCACAGATTG) GGGG A
pcDNA G2 miR392	(Rat SMN) From Mimoun Azzouz <u>Oligo sequences</u> Forward TGCTGTAAACCACGACACAGGTTTCTGTTTTGG CCACTGACTGACAGAAACCTGTCGTGGTTTA Reverse. CCTGTAAACCACGACAGGTTTCTGTCAGTCAGTGG CCAAAACAGAAACCTGTGTCGTGGTTTAC

Table 7.4. List of plasmids used in this thesis.

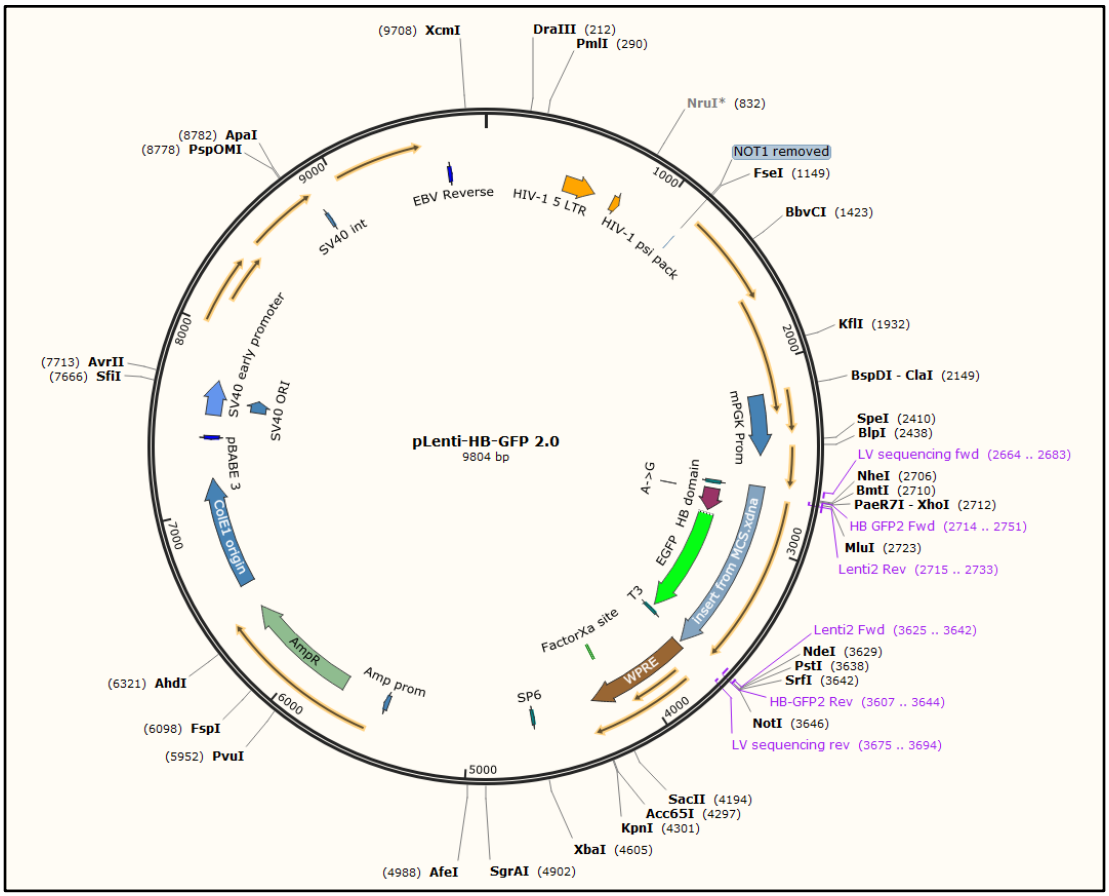

Figure 7.1. Graphic representation of pLenti-HB-GFP plasmid.

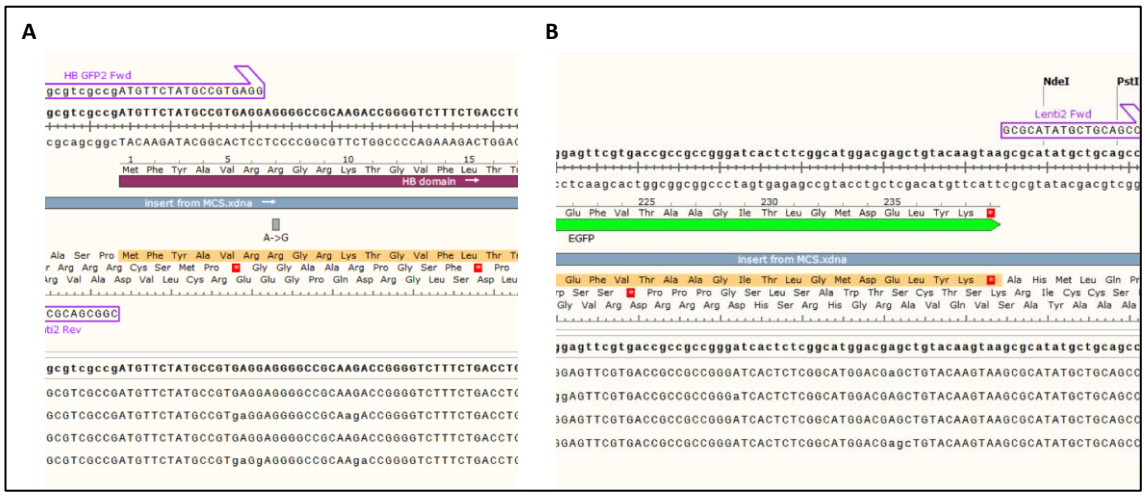

Figure 7.2. pLenti-HB-EGFP sequencing. Sequencing results from four pLenti-HB-EGFP plasmid clones, aligned with the sequence from the design map. (A) HB domain from the design in burgundy; sequencing results at the bottom. (B) EGFP sequence from the design in green; sequencing results at the bottom.

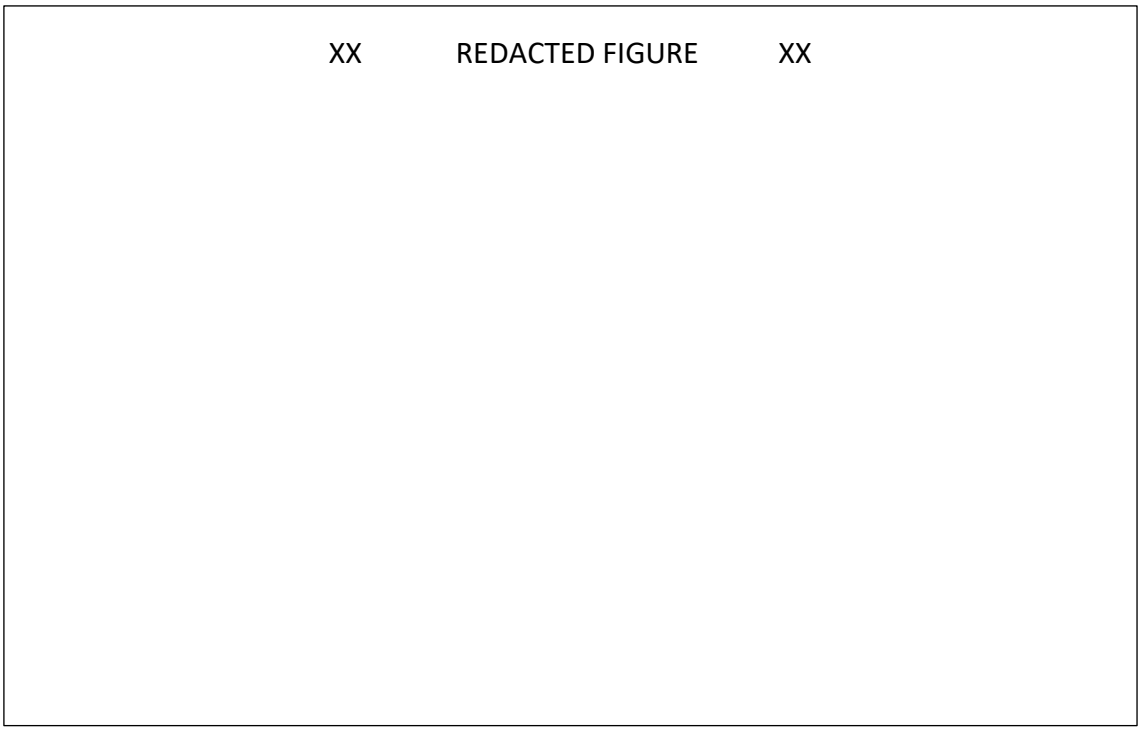

Figure 7.3. XX REDACTED TEXT XX.

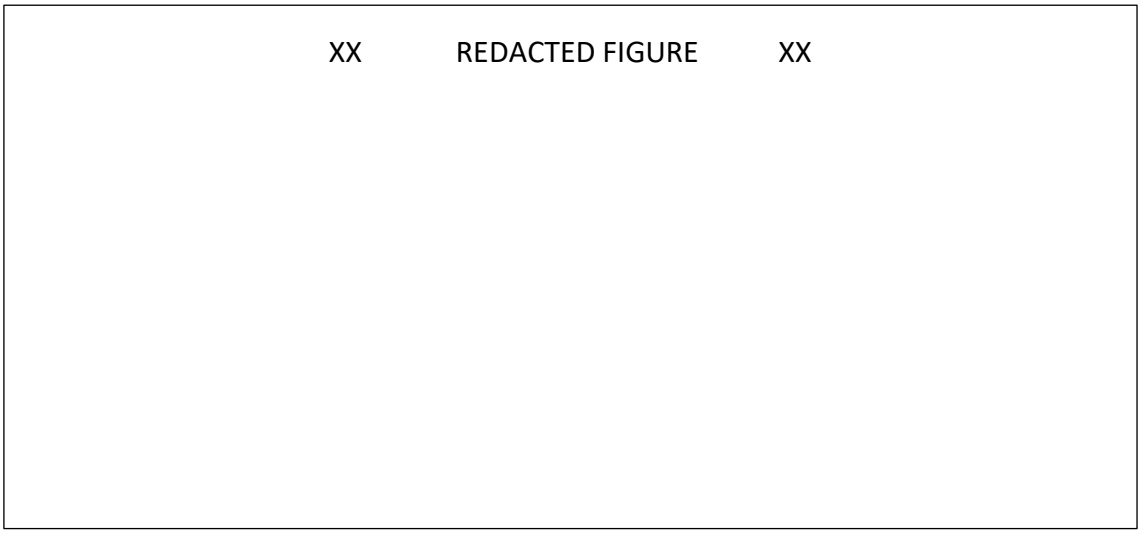

Figure 7.4. XX REDACTED TEXT XX.

XX REDACTED TABLE XX

XX REDACTED TABLE XX

CONTINUATION

XX REDACTED TABLE XX

CONTINUATION

Table 7.5. XX REDACTED TEXT XX.

XX REDACTED TABLE XX

XX REDACTED TABLE XX

CONTINUATION

XX REDACTED TABLE XX

CONTINUATION

XX REDACTED TABLE XX

CONTINUATION

Table 7.6. XX REDACTED TEXT XX.

XX REDACTED TABLE XX

XX REDACTED TABLE XX

CONTINUATION

XX REDACTED TABLE XX

CONTINUATION

XX REDACTED TABLE XX

CONTINUATION

XX REDACTED TABLE XX

CONTINUATION

XX REDACTED TABLE XX

CONTINUATION

XX REDACTED TABLE XX

CONTINUATION

Table 7.7. XX REDACTED TEXT XX.

XX REDACTED TABLE XX

XX REDACTED TABLE XX

CONTINUATION

XX REDACTED TABLE XX

CONTINUATION

Table 7.8. XX REDACTED TEXT XX

XX REDACTED TABLE XX

XX REDACTED TABLE XX

CONTINUATION

XX REDACTED TABLE XX

CONTINUATION

XX REDACTED TABLE XX

CONTINUATION

XX REDACTED TABLE XX

CONTINUATION

XX REDACTED TABLE XX

CONTINUATION

XX REDACTED TABLE XX

CONTINUATION

XX REDACTED TABLE XX

CONTINUATION

XX REDACTED TABLE XX

CONTINUATION

Table 7.9. XX REDACTED TEXT XX.