Decoding the role of the RNA binding activity of p53



Samantha Moore

Worcester College

University of Oxford

A thesis presented for the degree of Doctor of Philosophy

Michaelmas 2020

Contents

Abstract **7**

List of Figures **8**

List of Tables **11**

Abbreviations **13**

Acknowledgements **21**

1 Introduction **24**

1.1 Introduction 25

1.2 RNA binding proteins 25

1.2.1 Uncovering novel and unorthodox RNA binding proteins . . . 27

1.2.2 Identifying RNA binding domains 29

1.2.3 Unorthodox RNA binding domains: globular domains 30

1.2.4 Unorthodox RNA binding domains: intrinsically disordered regions 30

1.2.5 RNA binding proteins and binding specificity 31

1.3 RNA binding proteins and cancer 31

1.3.1 Cancer 31

1.3.2 RBPs and cancer . . . 32
1.4 The p53 tumour suppressor . . . 42
1.4.1 p53 protein structure . . . 43
1.4.2 p53 stabilisation and activation . . . 43
1.4.3 p53 as a transcription factor . . . 45
1.4.4 Additional functions of p53 . . . 46
1.5 p53-RNA interactions . . . 47
1.5.1 Anti-helicase & annealing activity . . . 48
1.5.2 Non-specific versus specific interactions . . . 49
1.5.3 Non-coding RNAs and p53 . . . 50
1.5.4 p53 binding interface . . . 51
1.6 Aims . . . 53

2 Materials **54**

2.1 Materials . . . 55
2.1.1 Chemicals, reagents, and consumables . . . 55
2.1.2 Buffers and solutions . . . 61
2.1.3 Commercial kits . . . 63
2.1.4 Antibodies . . . 64
2.1.5 Plasmids . . . 65
2.1.6 Oligonucleotides and primers . . . 66
2.1.7 Cell lines . . . 68

3 Methods **69**

3.1 Cell biology . . . 70
3.1.1 Cell culture . . . 70

3.1.2 Cell culture in SILAC media 70
3.1.3 Stable transfection of Flp-In T-REx cells 71
3.1.4 Platereader assay 71
3.2 Biochemical and molecular biology techniques 72
3.2.1 Cloning 72
3.2.2 Plasmid preperation 74
3.2.3 Western Blotting 75
3.2.4 Silver staining 75
3.2.5 RNA isolation, cDNA synthesis and RT-qPCR 76
3.3 RNase coupled proteomic analysis 76
3.3.1 RNase coupled immunoprecipitation 76
3.3.2 Mass spectrometry 78
3.4 RNA interactome capture 80
3.4.1 RNA interactome capture 80
3.4.2 Conventional protein analysis 81
3.4.3 Comparative RNA interactome capture 81
3.4.4 Mass spectrometry 82
3.5 CLIP-seq 84
3.6 iCLIP 86
3.6.1 Generation of irCLIP linker 86
3.6.2 iCLIP 87
3.7 Analysis 94
3.7.1 CLIP-seq analysis 94

4 Is p53 an RNA binding protein? 95

4.1 Introduction 96
4.2 p53 binds polyadenylated RNA in living cells 97
4.3 A lack of specificity of p53 antibodies presents challenges 100
4.4 An inducible GFP-p53 system increases specificity 104
4.5 eGFP-p53 interacts with known p53 interactors and RBPs 113
4.6 eGFP-p53 interacts with RNA in living cells 122
4.7 Discussion 125

5 Identifying the p53 RNA and protein interactomes 127

5.1 Introduction 128
5.2 Identifying the RNA targets of p53 129
5.3 Understanding p53-RNA binding at single nucleotide resolution . . . 143
5.4 The role of RNA in mediating p53 complexes 154
5.5 Discovering RNase sensitive p53-protein interactions 157
5.6 RNase-coupled proteomics reveals uncharacterised RBPs 163
5.7 RNA independent p53 protein interactions reveal DNA damage induced protein complexes 165
5.8 The RNA binding domain of p53 169
5.9 Discussion 172
5.9.1 p53 interacts with a defined subset of cellular mRNAs 172
5.9.2 Disparities in RNA-binding activity in genotoxic versus nongenotoxic p53 activation 177
5.9.3 Further analyses to reveal the non-coding RNA network of p53 179
5.9.4 iCLIP data will reveal p53's binding sites within target RNAs 180
5.9.5 RNA is a central mediator of p53 interactions 182

5.9.6 Prediction of p53's RNA binding domain 183

6 Global analysis of RBP responses to DNA damage 186

6.1 Introduction 187

6.2 Comparative RNA interactome capture allows comparison of RBPomes under different experimental conditions 189

6.3 Comparative RNA interactome capture yields high confidence RNA binding proteomes 192

6.4 Comparative RNA interactome capture highlights novel DNA damage associated RBPs 195

6.5 Comparative RNA interactome capture reveals changes to the RBPome after DNA damage 197

6.6 Discussion 201

6.6.1 Comparative RNA interactome capture reveals DNA damage induced changes to the RBPome 201

6.6.2 Future work 202

7 Discussion 204

7.1 p53 binds RNA in living cells 205

7.2 RNA plays a role in p53 protein interactions 206

7.3 p53 binds a distinct subset of mRNAs 207

7.4 DNA and RNA binding proteins 208

7.5 Post-transcriptional feedback loops 209

7.6 The RNA binding domain of p53 211

7.7 Summary and future perspectives 213

Appendices **214**

Abstract

Decoding the role of the RNA binding activity of p53

p53 is a key tumour suppressor, mutated in over half of cancers. p53's ability to form extensive protein interactions, along with its function as a sequence-specific transcription factor, are well documented due to their roles in canonical p53 tumour suppressor functions. Still lacking is clarity regarding p53's ribonucleic acid (RNA) binding activity. Previously reported, predominantly acellular and *in vitro* results provide little in regards to the physiological relevance, biological function and underlying mechanisms of p53-RNA interactions. Here I demonstrate p53 binds polyadenylated RNA in living cells using RNA interactome capture. Further investigation of p53-RNA interactions using conventional immunoglobulin G (IgG) antibody-based approaches proves problematic. I therefore generated an inducible cell line expressing green fluorescent protein conjugated p53 (GFP-p53), enabling highly specific purification of GFP-p53. Utilising CLIP-seq, I identified a defined subset of messenger RNAs (mRNAs) bound by p53, representing a translational programme distinct from p53's transcriptional targets. Additionally, RNase coupled proteomic analysis reveals the importance of RNA in p53's protein interaction network, including the DNA damage response. This work provides evidence of a p53 post-transcriptional regulatory layer, which warrants further investigation.

List of Figures

1.1 Cancer associated RNA binding proteins. 40

1.2 Network analysis of cancer associated RNA binding proteins. 40

1.3 Protein domains of unorthodox cancer-associated RNA binding proteins. 41

1.4 Modular structure of the p53 protein. 44

1.5 Summary of reported p53-RNA interactions. 47

4.1 RNA interactome capture protocol. 97

4.2 p53 binds polyadenylated RNA in living cells. 98

4.3 Immunoprecipitation using DO-1 antibody. 101

4.4 Immunoprecipitation using p53-trap nanobodies. 103

4.5 Proteomic analysis of p53-trap IPs. 105

4.6 Assessing GFP expression in GFP-p53 Flp-In T-REx cells. 106

4.7 Assessing GFP-p53 expression in GFP-p53 Flp-In T-REx cells. 108

4.8 Immunoprecipitation using GFP-trap nanobodies. 109

4.9 Assessing GFP-p53 induction upon DNA damage. 110

4.10 Immunoprecipitation using GFP-trap nanobodies. 112

4.11 Proteomic analysis of GFP-p53 IPs. 114

4.12 Proteomic analysis of GFP-p53 IPs. 115

4.13 Commonly co-immunoprecipitated proteins. 116

4.14 Nutlin/doxycycline - p53 protein interaction network. 118

4.15 Etoposide/doxycycline - p53 protein interaction network. 119

4.16 Proteomic analysis of GFP-p53 IPs reveals binding to RBPs. 121

4.17 Production of IR linker for labelling of protein-RNA complexes. . . . 123

4.18 Visualisation of GFP-p53-RNA complexes. 124

5.1 CLIP-seq protocol. 130

5.2 CLIP-seq UV crosslinking optimisation. 130

5.3 CLIP-seq analysis pipeline. 132

5.4 p53 binds mainly protein coding transcripts. 133

5.5 High confidence GFP-p53 bound transcripts. 138

5.6 PCA and hierarchical clustering of GFP-p53 bound targets. 140

5.7 Differential binding based on GFP-p53 induction condition. 141

5.8 iCLIP protocol. 145

5.9 Testing iCLIP immunoprecipitation conditions. 147

5.10 Optimising iCLIP partial RNase digestion. 149

5.11 Optimising iCLIP UV crosslinking conditions. 150

5.12 iCLIP test PCRs. 151

5.13 iCLIP iCount analysis pipeline. 153

5.14 RNase dependent protein-protein interaction protocol. 155

5.15 Nucleic acids impact MS processing. 156

5.16 Proteins significantly impacted by RNase treatment. 159

5.17 RNA binding proteins are impacted by RNase treatment. 160

5.18 RNase sensitive proteins - enrichment analysis. 162

5.19 Known p53 interactors are impacted by RNase treatment. 164

5.20 RNA independent p53 interactions. 166

5.21 RNA independent p53 interactions. 168

5.22 Prediction of p53's RNA binding sites. 171

6.1 Comparative RNA interactome capture protocol. 191

6.2 Silver staining of cRIC replicates. 192

6.3 DNA damage cRIC identifies known RNA binding proteins. 193

6.4 GO analysis of DNA damage cRIC identified proteins. 194

6.5 Domain architecture of DNA damage cRIC dataset. 196

6.6 GO analysis of candidate novel RBPs identified by cRIC. 197

6.7 6 hour etoposide cRIC. 198

6.8 24 hour etoposide cRIC. 199

6.9 Comparing 6 and 24 hour etoposide treatments using cRIC. 200

6.10 Comparison of DNA damage cRIC with etoposide RIC. 201

List of Tables

2.1 Reagents used in this study. 55

2.2 Buffers and solutions. 61

2.3 Commercial kits. 63

2.4 Antibodies. 64

2.5 Plasmids. 65

2.6 Oligos used for cloning. 66

2.7 Oligos used for RT-qPCR. 66

2.8 Oligos and primers used during the iCLIP protocol. 67

2.9 iCLIP Rclip primers. 67

2.10 Parental cell lines. 68

2.11 Self-generated modified cell lines. 68

3.1 Reaction mix for cloning PCR. 72

3.2 Cloning PCR cycling conditions. 72

3.3 Restriction enzymes used for p53 insert digestion. 73

3.4 Insert restriction digest. 73

3.5 Plasmid restriction digest. 73

3.6 Ligation reaction mixture. 74

3.7 Cycling conditions for RT-qPCR. 76

3.8 L3-IR linker 5' end pre-adenylation. 87

3.9 iCLIP PNK mix. 89

3.10 iCLIP ligation mix. 89

3.11 iCLIP reverse transcription reaction mix. 90

3.12 iCLIP reverse transcription PCR cycling conditions. 90

3.13 iCLIP ligation mix for circularisation. 91

3.14 iCLIP oligo annealing mix. 91

3.15 iCLIP oligo annealing conditions. 92

3.16 iCLIP test PCR mix. 92

3.17 iCLIP test PCR cycling conditions. 92

3.18 iCLIP final PCR mix. 93

5.1 Validating p53 *in vitro* interactors. 134

5.2 GFP-p53 + Nutlin/Doxycycline top candidate RNA targets. 135

5.3 GFP-p53 + Etoposide/Doxycycline top candidate RNA targets. . . . 136

5.4 Shared RNA targets. 137

5.5 iCLIP - final library PCR cycling conditions. 152

Abbreviations

A – Alanine

aa – Amino acid

ACN – Acetonitrile

AEBSF – 4-(2-aminoethyl)benzenesulfonyl fluoride hydrochloride

APS – Ammonium peroxide sulphate

ARE – AU-rich element

ARM – Arginine rich motifs

ATM – Ataxia telangiectasia mutated

ATP – Adenosine triphosphate

AUBP – AU binding protein

BSA – Bovine serum albumin

C- – Carboxy-

cDNA – Complementary DNA

ChIP-Seq – Chromatin immunoprecipitation and sequencing

CLIP – Crosslinking immunoprecipitation

CLIP-seq – Crosslinking immunoprecipitation and sequencing

Conc. – Concentration

COSMIC – Catalogue of Somatic Mutations in Cancer

CPEB4 – Cytoplasmic polyadenylation element-binding protein 4

cRIC – Comparative RNA interactome capture

CTD – C-terminal domain

D – Aspartic acid

DBCO – Dibenzocyclooctyne-amine

DBD – DNA binding domain

DBP – DNA binding protein

DMEM – Dulbecco's modified eagle medium

DMSO – Dimethyl sulphoxide

DNA – Deoxyribonucleic acid

DNA-PK – DNA dependent protein kinase

dNTP – Deoxyribonucleotide triphosphate

DRBPS – DNA and RNA binding proteins

DSS – Dextran sulfate sodium

DTT – Dithiothreitol

E – Glutamic acid

ECL – Enhanced chemiluminescence

EDTA – Ethylenediaminetetraacetic acid

eGFP – Enhanced GFP

eIF – Eukaryotic translation initiation factor

et al. – Et alia

ER – Endoplasmic reticulum

ERM – Ezrin, radixin, moesin

FBS – Foetal Bovine Serum

FGF-2 – Fibroblast growth factor 2

G – Glycine

GFP – Green fluorescent protein

GO – Gene Ontology

GSEA – Gene set enrichment analysis

HCL – Hydrochloric acid

HSP – Heat shock protein

HF – High fidelity

hnRNP – Heterogenous nuclear ribonucleoprotein

HRP – Horseradish peroxidase

hrs – Hours

HuR – Human antigen R

iCLIP – Individual nucleotide crosslinking and immunoprecipitation

IDR – Intrinsically disordered region

IgG – Immunoglobulin G

IP – Immunoprecipitation

IR – Ionising radiation

IRES – Internal ribosome entry sites

IRP1 – Iron regulatory protein 1

K – Lysine (Lys)

kDa – Kilodalton

KEGG – Kyoto encyclopedia of genes and genomes

KH – K Homology

L – Leucine

LB medium – Luria-broth medium

LC3 – Microtubule-associated proteins 1A/1B light chain 3

LDS – Lithium dodecyl sulfate

lincRNA – Long intergenic/intervening non-coding RNA

lncRNA – Long non-coding RNA

Lys – Lysine

mRNA – Messenger RNA

M – Methionine

MEF – Mouse embryonic fibroblast

Met – Methionine

mins – Minutes

miRNA – microRNA

MOPS – 3-(N-morpholino)propanesulfonic acid

mRNA – Messenger RNA

MS – Mass spectrometry

mTOR – Mammalian target of rapamycin

N- – Amino-/Amine-

NEAT1 – Nuclear Paraspeckle Assembly Transcript 1

NCI – National Cancer Institute

ncRNA – Non-coding RNA

NES – Nuclear export sequence

NLS – Nuclear localisation sequence

nm – Nanometer

nt – Nucleotide

nts – Nucleotides

OD – Oligomerisation domain

Oligos – Oligomers

OMIM – Online Mendelian Inheritance in Man

P – P-value or probability value

P – Proline

p53-RE – p53 response element

P/S – Penicillin/streptomycin

PAGE – Polyacrylamide gel electrophoresis

PEG – Polyethylene glycol

PBS – Phosphate-buffered saline

PBST – Phosphate-buffered saline + 0.1% Tween 20

PCR – Polymerase chain reaction

PK – Proteinase K

PNK – Polynucleotide kinase

Poly(A) – Polyadenylated/polyadenylation

PR – Proline-rich

PRR – Proline-rich region

PTBP1 – Polypyrimidine tract binding protein

PTEN – Phosphatase And Tensin Homolog

PTM – Post-translational modification

PUF – Pumilio and FBF

PVDF – Polyvinylidene fluoride

QKI – Quaking

qRT-PCR – Quantitative real time PCR

R – Arginine

RBD – RNA binding domain

RBM38 – RNA binding motif protein 38

RBP – RNA binding protein

RBPome – RNA binding proteome

REM – RNA, enzymes and metabolites

RFU – Relative fluorescent units

RIC – RNA interactome capture

RIPA – Radioimmunoprecipitation assay

RNA – Ribonucleic acid

RNP – Ribonucleoprotein complex

rpm – Revolutions per minute

RRM – RNA recognition motif

rRNA – Ribosomal RNA

ROS – Reactive Oxygen Species

RT – Room temperature

RT – Reverse transcription/transcriptase

RT-qPCR – Reverse transcription quantitative PCR

S – Serine

S.O.C – Super optimal broth with catabolite repression

SAM68 – Src-associated substrate during mitosis of 68 kDa

SDS – Sodium dodecyl sulphate

SDS-PAGE – Sodium dodecyl sulphate polyacrylamide gel electrophoresis

Seq – Sequencing

SF3B1 – Splicing Factor 3b Subunit 1

SFPQ – Splicing Factor Proline And Glutamine Rich

SILAC – Stable isotope labelling with amino acids in cell culture

snRNPs – Small nuclear ribonucleoproteins

SNV – Single nucleotide variant

SR – Serine- and arginine-rich

SRSF2 – Serine And Arginine Rich Splicing Factor 2

SRSF3 – Serine And Arginine Rich Splicing Factor 3

SRSF6 – Serine And Arginine Rich Splicing Factor 6

ss – Single stranded

ssDNA – Single stranded DNA

ssRNA – Single stranded RNA

STAR – Signal transduction activator of RNA metabolism

SUMO – Small ubiquitin-like modifier

T – Threonine

TAD – Transactivation domain

TAE – Tris base, acetic acid and EDTA

TBE – Tris-borate-EDTA

TBS – Tris-buffered saline

TBST – Tris-buffered saline + 0.1% Tween 20

TCA – Tricarboxylic acid cycle

TCEP – Tris(2-carboxyethyl)phosphine

TCR – T cell receptor

TE – Tris-EDTA

tRNA – Transfer RNA

U – Units

U2AF1 – U2 Small Nuclear RNA Auxiliary Factor 1

Ub – Ubiquitin

USER – Uracil-Specific Excision Reagent

UTR – Untranslated region

UV – Ultraviolet

VEGF – Vascular endothelial growth factor

WT – Wild type

WB – Western blotting

Y – Tyrosine

ZnF – Zinc finger

ZRSR2 – Zinc Finger CCCH-Type, RNA Binding Motif And Serine/Arginine Rich 2

α – Alpha

β – Beta

γ – Gamma

ϵ – Epsilon

γ – Gamma

^{o}**C** – Degree centigrade

g – RCF or G-force

hr – Hour

hrs – Hours

L – Litre

m – 10^{-3}

M – Molar

min – Minutes

O/N – Overnight

RT – Room temperature

sec – Second

secs – Seconds

μ – 10^{-6}

Acknowledgements

Firstly, thank you to Gareth Bond for giving me the opportunity to take on this project, and complete my DPhil.

Thank you to Alfredo Castello for your seemingly infinite enthusiasm, advice, patience and support. Thank you for giving me the opportunity to work in your lab - I have learnt more than I could have imagined. And thank you for always trying to get me to become a "glass half full" person.

Thank you to Ilan Davis for your advice, and never being too busy to help.

Thank you to the members of the Bond lab, past and present, for being so wonderful, and to all the members of the Ludwig Institute, who make it an inspiring place to work. Special thanks to Ping Zhang for all your help and advice, and to Giovanni Stracquadanio and Marsha Wallace.

Thank you to the members of the Davis lab, and the many many members of the Castello lab, past and present. You make coming to work everyday bearable. I am hugely grateful to Marko Noerenberg, Aino Järvelin, Wael Kamel and Manuel

Garcia Moreno for your advice and guidance, and never telling me you're too busy to help (most of the time).

Thank you to Worcester College for being home for the last 9 years.

Thank you to Aino Järvelin, Colin Goding and Clive Moore for your help in the preparation of this manuscript.

Thank you to my beautiful friends. Thank you for putting up with me when I go missing, and for always being there for me.

Thank you to my Mom and Dad, for constantly supporting me and pushing me. I love you both very much.

The most important thank you is to my husband, Sam. I would not have been able to do any of this without you. I couldn't wish for a more supportive, loving and understanding partner. I love you, and thank you.

1

Introduction

1.1 Introduction

Eukaryotic organisms' ability to regulate gene expression is essential for development, differentiation, cellular homeostasis and response to both intrinsic and extrinsic stressors. Transcriptional regulation of gene expression has been extensively studied over the past three decades, highlighting the importance of transcription factors and histone and chromatin modifying factors in transcriptional modulation. However, investigation of co-transcriptional and post-transcriptional regulatory layers lagged behind. Recent technological advances have enabled exploration of the role of co- and post-transcriptional processes in enabling the regulation and fine-tuning of eukaryotic gene expression. Accordingly, this has led to increased appreciation of the importance of these processes in normal cellular function, along with their dysregulation in a range of diseases.

1.2 RNA binding proteins

RNA-binding proteins (RBPs) are crucial co-ordinators of gene expression. RBPs bind ribonucleic acid (RNA) molecules throughout their lifecycle, forming highly dynamic ribonucleoprotein complexes (RNPs). These RNPs are crucial in the regulation of RNA fate, and every stage of the RNA production and maturation, including transcription, processing (splicing, capping, polyadenylation), export, subcellular localisation, storage, translation and decay. RNPs are incredibly dynamic complexes, undergoing remodelling throughout the RNA lifecycle, but also in order to co-ordinate gene expression programmes in response to extracellular and intracellular stimuli. In accordance with this, RNA-binding proteomes (RBPomes) have

been shown to be highly cell specific, and to undergo extensive remodelling in response to a range of intracellular and extracellular signals, including viral infection [1], arsenite treatment [2], hypoxia [3] and exposure to ionising radiation (IR) [4].

A single RBP can bind many diverse RNA targets. Conversely, an individual RNA is in complex with multiple RBPs throughout its lifetime; often concurrently and in a highly dynamic manner. Binding of multiple RBPs to RNA targets can be synergistic, or co-operative, through direct protein interaction or via a shared RNA target. Alternatively, binding can be antagonistic or competitive, involving one or more RNA binding sites, with the balance of competitive RBPs ultimately defining the regulatory outcome. Additionally, RNP complexes contain additional proteins in complex with RNA-bound RBPs, further increasing the complexity of posttranscriptional regulatory networks.

Conventional perspectives on RNA-protein interactions view the RBP as the protagonist; binding the RNA and enforcing its impact upon the RNA in some way. This is evident in the literature that, until recently, focussed predominantly on messenger RNA (mRNA) and the way in which RBPs were able to impact upon mRNA transcription, processing, export, translation and decay; ultimately impacting gene expression. However, it is becoming more apparent that RNA-centric views of RNA-protein interactions are also valid, and in some cases more biologically appropriate. RNAs can bind to RBPs, and impact RBP localisation, stability and function, along with proteins' ability to interact with other RNA or protein species. Indeed, RNAs can be viewed in many cases as "scaffolds" upon which protein complexes can be coordinated, or sequestered as is the case with the NEAT1 (Nuclear Paraspeckle Assembly Transcript 1) long non-coding RNA (lncRNA), which sequesters RBPs

into paraspeckles [5]. Protein-RNA associations can be thought of as occupying a spectrum between protein-determined and RNA-determined interactions [6].

Conventionally, RBPs are modular proteins, binding target RNAs using a combination of well-defined RNA binding domains (RBDs), such as the RNA recognition motif (RRM), K Homology (KH) domain, Piwi and Pumilio and FBF homology (PUF) domains. RBDs generally recognise stretches of RNA 2-10 nucleotides (nts) long [7], with relatively low affinity. However, through combinatorial and concurrent binding of multiple modular domains, including auxiliary domains, RBPs are able to bind an expansive range of RNA targets with high affinity, specificity and versatility.

1.2.1 Uncovering novel and unorthodox RNA binding proteins

In the quest to identify unknown RBPs, *in silico* approaches to predict proteins with RNA binding capacity, based on the presence of well characterised RBDs, allowed further expansion of the RNA binding proteome. However, these approaches are limited to proteins that bind RNA via well characterised RBDs. *in vitro* techniques to comprehensively identify RNA binding proteins, such as the use of protein arrays, further expanded the repertoire of RNA binding proteins, and continued to reveal RBPs with no previous link to RNA biology, such as metabolic enzymes [8].

Methodological advances over the last decade have enabled the development of techniques that allow investigators to comprehensively identify the RNA binding proteomes of living cells, including RNA interactome capture (RIC [9, 10]). RIC utilises UV irradiation to induce the formation of covalent bonds between natu-

rally photoreactive nucleotide bases and amino acids, essentially freezing RNA and proteins in direct contact (termed zero distance). Purification of polyadenylated (poly(A)) RNA-bound proteins via oligo(dT) capture, and subsequent proteomic analysis, enables identification of native RBPomes. The RBPomes of several human and mouse cell types, along with those of several model organisms, such as *Saccharomyces cerevisiae*, *Caenorhabditis elegans* [11] and *Arabidopsis thaliana* [12] have recently been elucidated. Comparison of eukaryotic RBPomes reveal a high level of conservation, and a set of approximately 250 core eukaryotic RBPs [13]. To date, between 1500 and 2000 human RBPs have been catalogued [11].

During this influx of RNA binding proteomic data, it became apparent that many of the identified novel RBPs lacked the well-defined RBDs characteristic of orthodox RBPs [14]. Approximately 55% of currently identified human RBPs lack a classical RBD or other characterised RNA binding interface [15]. These "unorthodox" RBPs comprise a wide range of proteins, often with functions previously unrelated to RNA biology, including a large number of metabolic enzymes, involved in carbohydrate, lipid, amino acid and nucleotide metabolism [8]. The prevalence of "moonlighting" enzymes as novel RBPs suggests links between RNA biology and metabolism, as suggested by the REM (RNA, enzymes and metabolites) network hypothesis, which postulates a link between metabolism and the regulation of gene expression via these enzymes with RNA binding functions [16]. Additionally, proteins that have previously been characterised as DNA (deoxyribonucleic acid) binding are increasingly identified as possessing RNA binding activities [17]. Not only do unorthodox RBPs pose questions regarding the biological functions of previously unidentified RBPs in RNA metabolism, but also regarding the mechanisms and modality of binding in

the absence of characterised RBDs.

1.2.2 Identifying RNA binding domains

As the number of RNA binding proteins has expanded, so too have attempts to identify the protein interfaces responsible for RNA binding. *in vitro* approaches utilising mass spectrometric identification of RNA-bound peptide fragments have enabled the identification of RNA binding domains of a few hundred proteins [15]. Additionally, computational approaches have facilitated identification of well-defined RBDs across the proteome.

RBDmap [15] utilises successive rounds of oligo(dT) capture, interspersed with proteolytic digestion of RNA bound proteins, to enrich for protein peptides engaged in RNA binding in living cells. Peptides are then subject to tryptic digestion and quantitative proteomics. The nucleotide(s) crosslinked to the peptide makes identifying bound peptides extremely challenging, due to aberrant and heterogeneous mass shift [15]. However, as the first protease cleaves less frequently than trypsin, the latter would liberate a nucleotide-crosslinked peptide and likely, an unmodified tryptic peptide. Identification of these unmodified peptides are used as tags to reconstitute, *in silico*, the original peptide in contact with the RNA. RBDmap has enabled the identification of RNA binding interfaces on a proteome-wide scale; not only validating the utility of well-defined globular RBDs in RNA-protein interactions, but also identifying the RNA binding interfaces of many previously uncharacterised, unorthodox RBPS.

1.2.3 Unorthodox RNA binding domains: globular domains

In many cases, unorthodox RBPs are demonstrated to bind via unorthodox globular domains [6]. These can include dual functional domains, including enzymatic cores, DNA binding domains, protein interaction interfaces, and domains involved in protein localisation. These dual functioning domains may represent a mechanism by which RNA binding can be modulated in accordance with the RBPs' other function, for example the binding of metabolites may impact RNA binding, as is the case with the iron regulatory protein IRP1 [18].

1.2.4 Unorthodox RNA binding domains: intrinsically disordered regions

Intrinsically disordered regions (IDRs), defined as peptide regions lacking a stable tertiary structure in solution [7, 19], are present in both unorthodox and conventional RBPs. Approximately 20% of known mammalian RBPs contain over 80% sequence disorder [7, 19], whilst over 50% of all peptides identified by RBDmap aligned to disordered regions [15]. IDRs can combine in a modular manner with orthodox RBDs, or can comprise a protein's sole RNA interaction interface. Typically IDRs are composed of disorder promoting amino acids, such as serine (S), glycine (G) and proline (P), along with charged amino acids such as lysine (K), arginine (R), aspartate (D) and glutamate (E). IDRs can bind via modular repeats, such as RS, RG and arginine rich motifs (ARMs), low complexity regions, as well as basic patches. Additionally, basic tails similar to analogous sequences in DNA binding proteins (DBPs) can facilitate RNA binding; providing a potential link between RNA and DNA binding proteins. The general flexibility of IDRs means they are able to facilitate multiple

interaction partners, often through disorder to order transitions upon folding [20]. IDRs are generally hotspots for post translational modifications (PTMs), which can represent a mechanism by which RNA binding can be modulated, either through alteration of binding properties, subcellular localisation or protein function.

1.2.5 RNA binding proteins and binding specificity

RNA-protein interactions fall along a continuum, with highly specific interactions mediated by nucleotide and shape complimentarity at one end, and promiscuous RNA binding by interaction with the RNA sugar phosphate backbone at the other. Within these two extremes lies a spectrum of specificities, including recognition of the RNA molecule as a whole, and of RNA secondary or tertiary structures [7]. Binding specificities are only known for a small subset of extremely well characterised RBPs, such as the PUF family of proteins [6]. It is worth acknowledging that, in the case of RBP-RNA interactions, neither high specificity nor high affinity automatically equates to functionality. For example, the eukaryotic translation initiation factor (eIF) 4F enables translation initiation through sequence-independent interaction with the mRNA 5' cap [11]

1.3 RNA binding proteins and cancer

1.3.1 Cancer

Cancer includes a range of heterogenous diseases, resulting from the stepwise accumulation of genetic and/or epigenetic alterations that disrupt normal cellular mechanisms regulating cell proliferation. During carcinogenesis, the normal balance

between proto-oncogene signalling and tumour suppressor activities is disrupted, resulting in carcinogenesis and neoplastic growth. Hanahan and Weinberg [21] described the hallmarks of cancer, which detail the six fundamental alterations to normal cell physiology that enable neoplastic growth and metastasis. These "hallmarks of cancer" are: self sufficiency in growth signals, insensitivity to anti-growth signals, evasion of apoptosis, sustained angiogenesis, limitless replicative potential, and tissue invasion and metastases. With subsequent progress in our understanding of cancer, these hallmarks were expanded to include: evasion of immune destruction, deregulation of cellular energetics, genome instability and mutation, and tumour promoting inflammation [22].

1.3.2 RBPs and cancer

Given the pivotal role of RBPs in post transcriptional regulation and RNA fate, it is unsurprising that disruption of normal RBP function can result in disease. Indeed, dysregulation of a host of RBPs, via several different mechanisms, have been observed across a range of cancer types; highlighting the importance of post-transcriptional regulatory networks in neoplastic disease [23, 24]. These post-transcriptional regulatory networks are complex. As a result, even minor deviations from normal RBP activity or function can have significant downstream impacts, influencing a range of pathways. Conversely, cancers are also able to utilise and exploit post-transcriptional regulatory molecules, such as RBPs and microRNAs (miRNAs), using them to their advantage. As such, RBPs can influence multiple cancer-associated pathways and traits, and are usually associated with multiple cancer hallmarks [25].

As RBPs play pivotal roles in a broad range of co-transcriptional and post-

transcriptional processes, the impacts of RBP function and dysfunction in cancer are equally as broad. The dysregulation of RBPs in the context of neoplastic disease is therefore complex, involving interconnected layers of regulation, not purely reliant on genomic alterations to the RBP itself, and often impacting on a range of downstream cancer associated processes and pathways.

RBP dysregulation in cancer

Despite genome instability and mutation being an emerging hallmark of cancer, genomic alterations appear to play a relatively minor role in the dysregulation of RBPs. Copy number variations [25], along with epigenetic modification, represent prevalent modes of RBP dysregulation, enabling both overexpression and downregulation of target RBPs, although consensus over which is more commonly observed is debated. Within the context of cancer, 15% of RBPs have mutations to their protein coding sequences [25]. Germ-line mutations and chromosomal translocations are both rare, with approximately 1% of RBPs impacted by germline mutations [25], and few examples of chromosomal translocations/fusion proteins [25, 26]. Evidence to date suggests alterations to RBPs are generally not driver mutations, and that RBP's impacts on cancer are normally subsequent to original oncogenic events [25]. In addition to genetic variations or lesions impacting RBPs, there are several other potential mechanisms of RBP dysregulation in cancer.

Cancer drivers – genes in which mutations provide selective advantage to cancer – have been shown to impact RBPs *in vitro*. For example, in cancer cell lines, the p53 tumour suppressor protein can induce RNA binding motif protein 38 (RBM38), which enables stabilisation of target mRNAs, such as *p21*, which in turn promotes G1 cell cycle arrest [25].

RBPs are subject to post-transcriptional regulation by other RBPs, as well as non-coding RNAs (ncRNAs) such as miRNAs. Global reduction in miRNA is a feature of cancers [25], and as such may perturb normal RBP regulation. Conversely, RBPs may also impact miRNA synthesis and regulation, resulting in a feedback loop. For example, DICER1, which itself is involved in the maturation of miRNAs, is negatively repressed by *miR-103* and *miR-107*, which in turn leads to disrupted miRNA biosynthesis, and metastasis in breast cancer cell lines, as well as in a specific breast cancer cohort [25].

As previously mentioned, post translational modification of RBPs represents an additional mechanism by which RBP activity can be modulated. Acetylation levels have been observed to be higher in cancer than in normal cells. Src-associated substrate during mitosis of 68 kDa (SAM68) acetylation is enhanced in cancer cells, leading to increased RNA binding, impacting upon regulation of the cell cycle. Similar patterns are seen for phosphorylation of RBPs [27].

To date, our understanding of the precise molecular mechanisms underpinning the impacts of RBPs in cancer is fragmented. Below I list a subset of the ways in which altered or aberrant RBP functions contribute to cancer traits.

RBPs involved in splicing and cancer

RNA splicing is key for processing of pre-mRNAs, with its function being to remove non-coding introns. Splicing is performed by the spliceosome, comprising of small nuclear ribonucleoprotein (snRNP) complexes, which are made up of short non-coding RNAs, an Sm or Sm-like protein, and additional snRNP specific proteins. Normal splicing is regulated by both *cis*-regulatory elements, including the 5' and 3' splice sites, and *trans*-acting factors. The predominant *trans*-acting factors are the

serine- and arginine-rich (SR) and heterogenous nuclear ribonucleoprotein (hnRNP) proteins, but other proteins are involved in splicing regulation. Alternative splicing enables cells to generate different protein isoforms from the same gene. Dysregulation of both constitutive and alternative splicing appears to be common across a range of cancers, through impacts on both *cis* and *trans*-acting factors.

Genes encoding core spliceosomal proteins have been observed to be subject to recurrent somatic mutations, with 4 genes specifically targetted: *SF3B1* (Splicing Factor 3b Subunit 1), *SRSF2* (Serine And Arginine Rich Splicing Factor 2), *U2AF1* (U2 Small Nuclear RNA Auxiliary Factor 1) and *ZRSR2* (Zinc Finger CCCH-Type, RNA Binding Motif And Serine/Arginine Rich 2), which are all involved in 3' splice site recognition [28]. These mutations appear to be mutually exclusive, and occur in a consistently heterozygous state, suggesting the need for a level of wild type splicing function . These mutations were initially identified in haematological malignancies, but have since been found in other solid tumour types [28].

Even in the absence of mutation, several splicing factors, including hnRNP and SR proteins, exhibit altered expression or activity, which can contribute to carcinogenesis and altered response to treatment [29, 28]. For example, quaking (QKI) is a STAR (signal transduction activator of RNA metabolism) family protein, that is frequently downregulated in lung cancers. Normally, QKI appears to act as a tumour suppressor, by promoting expression of a NUMB isoform that negatively regulates the Notch signalling pathway, through repression of exon 12 inclusion [28]. Other factors, such as SRSF3 and SRSF6 (Serine And Arginine Rich Splicing Factors 3 and 6) have been described as proto-oncogenes, showing overexpression in a range of cancers. Recurrent translocations can also be seen for a selection of splicing factors,

including *SRSF3* and *SFPQ* (Splicing Factor Proline And Glutamine Rich) [30, 31].

Cis-regulatory elements are also susceptible to alteration in cancers. For example somatic single nucleotide variants (SNVs) were observed to affect splice sites of several tumour suppressors, including the p53 encoding *TP53* and *PTEN* (Phosphatase And Tensin Homolog), which resulted in intron retention, and subsequently premature stop codons [28].

RBPs involved in alternative polyadenylation and cancer

Addition of the 3' poly(A) tail is a key mRNA processing step, that involves recognition of a poly(A) site, followed by endonucleolytic cleavage of mRNA transcripts, and addition of a string of adenosines. The poly(A) tail plays a key role in mRNA stability, nuclear export, and ultimately translation. Alternative polyadenylation involves utilisation of alternative poly(A) signals on the pre-mRNA, resulting in different length isoforms, differing in either their 3' untranslated region (UTR) or coding sequence.

Cancer cells demonstrate widespread shortening of 3' UTRs, which is thought to enable reduced binding of regulatory *trans*-acting factors, including RBPs and ncRNAs [27]. This global phenomenon is thought to be mediated, in part, by subunits of the cleavage factor Im complex, which mediates recruitment of protein complexes to the cleavage reaction [32]. Interestingly, the alternative polyadenylation of specific transcripts can also be deregulated in cancer. For example the cytoplasmic polyadenylation element-binding protein 4 (CPEB4) is upregulated in a range of cancers, including melanoma. Within the context of melanoma, CPEB4 regulates polyadenylation, and subsequent translational upregulation, of the melanoma drivers RAB72A and MITF, which in turn promote proliferation [25].

RBPs involved in mRNA stability and cancer

AU-rich elements, or AREs, are sequences found within the 3' UTR of approximately 16% of human transcripts [25]. These ARE containing transcripts are liable to decay, and binding by ARE-binding proteins (AUBPs) can either facilitate or prevent this. ARE-containing mRNAs are overrepresented in multiple cancer types, especially in oncogenes, cell cycle regulators and growth factors [33]. As such, dysregulation of AUBPs can impact upon the stability of their target transcripts, which have clear links to carcinogenesis. The human antigen R protein (HuR), for example, is upregulated in a range of cancers, and is termed "the regulator of regulators", associated with every hallmark of cancer. HuR is able to enhance the stability of several ARE-containing transcripts encoding regulators of the cell cycle, such as cyclin E1, and thus enables increased proliferation [27, 25].

RBPs involved in translation and cancer

Dysregulation of translation is a common feature of cancers. Hyperproliferating cancer cells require increased protein synthesis to support their extreme cell growth, which is generally achieved through hyperactivation of the translational machinery [34], along with increase in ribosome biogenesis [35]. Along with a global increase in protein synthesis, the translational upregulation of specific genes that facilitate tumourigenesis and cancer cell survival is also observed [36].

Translation initiation factors are frequently dysregulated in cancer, which can lead to global, as well as specific alterations to translation. The role of the eIF4E component of the cap binding complex is the best studied example of this. Over-expression of eIF4E, either by gene amplification or as a result of hyperactivation

of the mammalian target of rapamycin complex (mTORC), results in the selective translational upregulation of a range of pro-tumorigenic mRNAs, including *VEG-F* (vascular endothelial growth factor), *FGF-2* (fibroblast growth factor 2) and *MDM2* [34, 37].

Mutations impacting ribosomal components, and factors involved in ribosome biogenesis, are causally linked to a range of syndromes termed 'ribosomopathies' and accompanying predisposition to cancer [38].

Along with alterations to the translational machinery, oncogenic translational programmes are significantly impacted by 3' and 5' *cis*-regulatory elements [27, 39]. For example, the internal ribosome entry site (IRES) enables cap-independent translation initiation, which can be utilised when conventional cap-dependent translation is repressed, for example during oncogenic stress. The HuR RBP is shown to modulate IRES-dependent translation of several mRNAs, including X-linked inhibitor of apoptosis (*XIAP*) [27, 25]. Interestingly, HuR is shown to regulate a range of anti-apoptotic targets. This is an example of an RNA operon, in which an RBP, in conjunction with ncRNAs, is able to modulate the expression of a specific subset of mRNAs with shared function and/or subcellular location [40].

This is not an exhaustive review of the mechanisms of RBP action in cancer, with further roles for RBPs in other RNA metabolic processes such as RNA editing, and subcellular localisation. What is demonstrated, is the range of mechanisms through which RBP function and dysfunction are able to impact cancer.

Unorthodox RBPs and cancer

As technological advances have enabled us to expand the repertoire of known cellular RBPs, the number of unorthodox RBPs without previous links to RNA biology

has also increased. Unavoidably, as a result there are an increasing number of unorthodox RBPs associated with disease, including cancers. The role of these unorthodox RBPs in cancer remain unclear, but will become more apparent as the RNA binding activities of these uncharacterised unorthodox RBPs are studied. We recently reviewed the role of unorthodox RBPs in cancer [41].

RNA interactome data from a range of cell types [11], and RBDmap data [15] were combined to provide a superset of identified RNA binding proteins, with accompanying binding site information. Identified RBPs were then overlapped with cancer associated proteins obtained from the Catalogue of Somatic Mutations in Cancer (COSMIC) [42], Online Mendelian Inheritance in Man (OMIM)[43] and a curated list of cancer-associated Gene Ontology (GO) terms [44].

In this way, 696 RBPs were identified as cancer-associated (Figure 1.1); representing almost half of the RNA-bound proteome, and highlighting the importance of RBPs in cell fate decisions. Interestingly, the distribution of cancer-associated RBPs was almost equal between classical and unorthodox RBPs (40.9% classical RBPs, 59.1% unorthodox).

Network analysis of the cancer-associated RBPs reveals clear functional clusters (Figure 1.2), Prominent clusters include ribosome/ribosomal biogenesis, splicing and transcription, which is perhaps unsurprising given cancer's propensity to exploit post-transcriptional processes, as described previously. Other potentially expected groups of RBPs include those involved in RNA localisation, RNA degradation and transfer RNA (tRNA) biosynthesis. Other functional complexes identified by network analysis highlight interesting links between RNA biology, RBPs and cancer, with links to the emerging hallmarks of cancer.

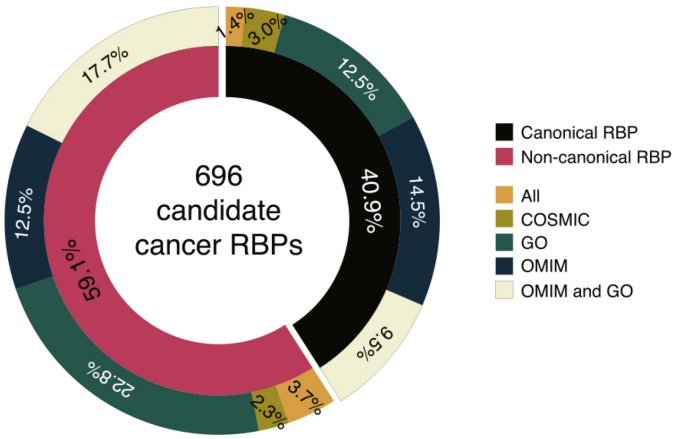


Figure 1.1: Cancer associated RBPs. Sunburst graph showing breakdown of cancer-linked RNA binding proteins (RBPs) by RNA binding domain classification (i.e. classical or non-canonical) and source of cancer association information. Generated by Aino Järvelin.

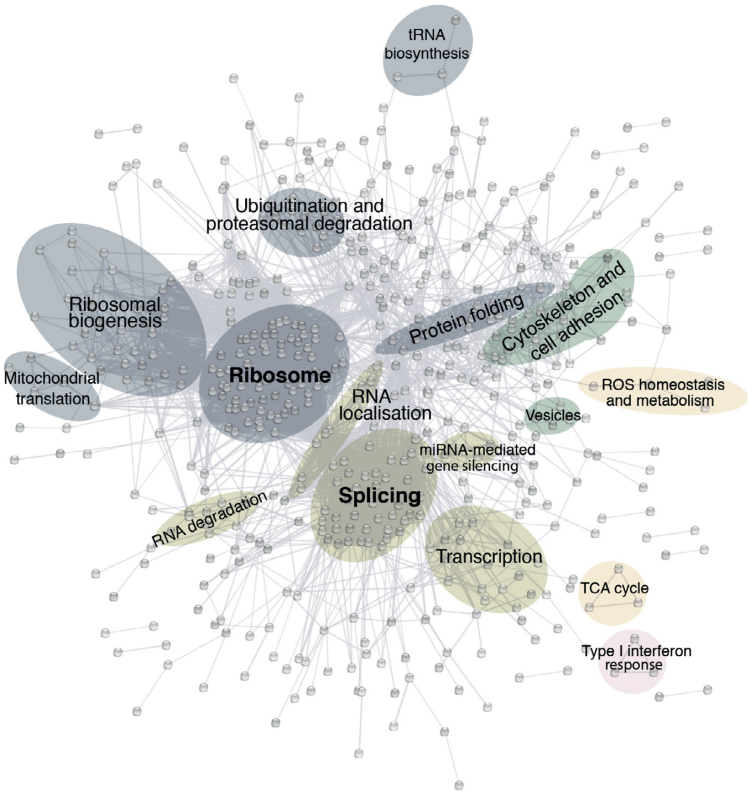


Figure 1.2: Network analysis of cancer associated RBPs. STRING [61] network showing connections between RNA binding proteins with links to cancer based on annotations in COSMIC, OMIM, and GO databases. Functional complexes are annotated.

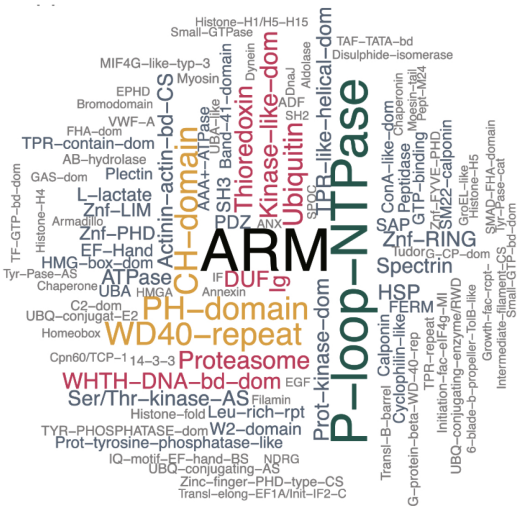


Figure 1.3: Protein domains of unorthodox cancer-associated RBPs. Word cloud showing domains present in unorthodox cancer-linked RNA binding proteins. Relative prevalence is indicated by size. Generated by Aino Järvelin.

Dysregulation of cellular metabolism is an emerging hallmark of cancer [22], pivotal in facilitating the increased demand on cells undergoing neoplastic growth. Identification of moonlighting enzymes as novel RBPs provides a link between RNA regulation and metabolism. Interestingly, two enzymatic clusters are highlighted by network analysis; enzymes involved in the tricarboxylic acid cycle (TCA cycle) and a group of enzymes involved in homeostasis and metabolism of reactive oxygen species (Figure 1.2), linking antioxidant metabolism, RNA and cancer. Additionally, enzymatic cores represent predominant domains in unorthodox RBPs (Figure 1.3).

Neoplastic cells' capacity to invade surrounding cells and tissue, and ultimately metastasise, is a key hallmark of cancer [21]. All three components of the ezrin,radixin,moesin (ERM) family are identified as cancer-associated RBPs. The ERM complex plays key roles in signal transduction, as well as organisation of cell junctions, adhesion sites and specialised membrane structures, through mediation of interactions between cytoskeletal elements and the plasma membrane. These functions have pre-

viously been linked to cancer progression and metastasis, through regulation of cell adhesion and migration [45].

Additionally, a number of heat shock proteins (HSPs) are identified, which act as molecular chaperones, facilitating protein synthesis, localisation and degradation. HSPs confer cytoprotective effects to cells under cellular stress, thus facilitating cellular survival. Overexpression of HSPs is observed in a range of cancers, which monopolise on the chaperone's anti-apoptotic and pro-survival functions to facilitate tumour growth [46]. Indeed, the utility of HSP inhibitors is being studied, as HSPs client proteins include several members of oncogenic signalling pathways [47].

As we continue to identify novel RBPs, our understanding of the complex links between RNA biology and other cellular processes will also continue to expand.

1.4 The p53 tumour suppressor

The p53 tumour suppressor protein was discovered in the 1970s by David Lane and Lionel Crowford [48], and Arnold Levine and Daniel Linzer [49], and has since become one of the most extensively studied proteins [50]. The *TP53* gene, encoding the p53 protein, was initially described as an oncogene, capable of immortalising cells either alone or in conjunction with Ras [50]. However, continued study of p53 revealed that wild type p53 could actually suppress cellular transformation, whilst p53 possessing point mutations was indeed transforming, thus establishing p53 as a tumour suppressor [51]. The importance of p53's tumour suppressor functions in preventing neoplastic transformation and tumour progression is reflected by the fact that the *TP53* gene is mutated in over half of sporadic cancers, whilst germline mutations result in the highly penetrant Li-Fraumeni cancer predisposition syndrome.

1.4.1 p53 protein structure

p63 and p73 comprise the other members of the highly conserved p53 protein family. Wild-type p53, encoded by 11 exons, is a protein of 393 amino acids, and possesses a modular structure (Figure 1.4). Two N-terminal transactivation domains (TAD) enhance p53's transcriptional activity, through the recruitment of coactivators, histone modifying enzymes and components of the basal transcription machinery. A proline rich region (PRR) lies C-terminal to TAD1 and TAD2, which contains multiple PXXP repeats, and is thought to mediate protein-protein interactions. The core DNA binding domain (DBD) is responsible for sequence specific binding of p53 to p53 response elements (p53-REs), and represents the most highly conserved region of the protein. Proximal to the DBD lies the hinge domain, and the oligomerisation domain (OD), also referred to as the tetramerisation domain, which facilitates the formation of p53 oligomers, including the functional p53 tetramer. Finally, there is the C-terminal domain (CTD). This intrinsically disordered domain is subject to extensive post-translational modifications, which are crucial in the regulation of p53 stability, activity and function. Despite the C-terminal domain often being described as a regulatory domain, it is involved in nearly every aspect of p53 function [52].

1.4.2 p53 stabilisation and activation

Under normal physiological conditions, p53 is kept at low levels, with a short half life (6-20 minutes) [53]. This predominantly occurs through an autoregulatory feedback loop with its main regulator MDM2. MDM2 is an E3 ubiquitin ligase, which binds to p53's TAD. MDM2 ubiquitinates p53, targeting it for proteasomal degradation,

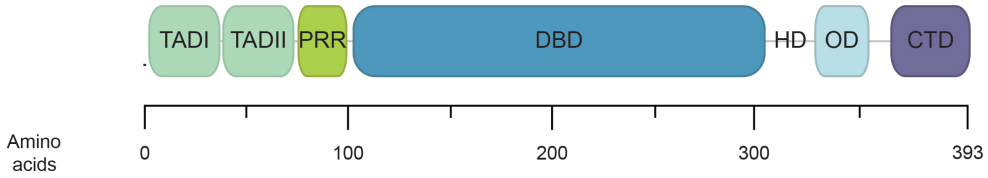


Figure 1.4: Modular structure of the p53 protein. Schematic representation of the p53 protein, with functional domains annotated. Transactivation domain (TAD), proline rich region (PRR), DNA binding domain (DBD), hinge domain (HD), oligomerisation domain (OD), C-terminal domain (CTD)

whilst also blocking its ability to recruit transcriptional machinery. In turn, p53 activates *MDM2* transcription. Additional E3 ubiquitin ligases are involved in the regulation of p53 levels under physiological conditions, including PIRH2, COP1 and ARF-BP1. Another ubiquitin ligase involved in the negative regulation of p53 is the MDM2-family member MDM4, which prevents p53's transcriptional activities, and appears to facilitate MDM2-mediated p53 regulation [54].

Upon stress, the MDM2-p53 interation is inhibited through post-translational modifications, including phosphorylation of Ser395 and Tyr394 on MDM2 by ATM (Ataxia telangiectasia mutated) and c-Abl kinases. Additional modifications of p53 can facilitate the disruption of this interaction, by stress-induced kinases such as ATM, ATR, Serine/threonine-protein kinase Chk1 and Chk2, and DNA-dependent protein kinase (DNA-PK). Disruption of the MDM2-p53 interaction prevents p53 degradation, and consequently enables stabilisation and accumulation of p53 within the cell. An array of subsequent PTMs are thought to further finetune p53 activity, and thereby modulate p53's response to cellular stress [55, 56].

1.4.3 p53 as a transcription factor

p53 is a transcription factor, whose regulation of a myriad of transcriptional targets enables it to mount responses to cellular stress. p53 recognises DNA sequences, termed p53-REs, which are defined as two half sites comprising the motif RRRCWWGYYY (where R dictates a purine, W an adenine or thymine and Y a pyrimidine), separated by a 0-13 base pair spacer [57]. p53 is able to bind a vast number of sites across the genome, thus regulating the transcription of a range of protein coding and noncoding transcripts.

Through its function as a sequence specific transcription factor, p53 is able to regulate the expression of transcripts, which enable it to execute a range of stress-response programmes, depending on the nature of the cellular stress, environmental factors, and cell type. Under extreme stress, p53 is able to trigger apoptosis through induction of both intrinsic and extrinsic pathways, including PUMA, BAX, NOXA and FUS [58, 59]. p53 can also induce cellular senescence via expression of p21 and PML, amongst others. Under less severe stress, where p53 can facilitate repair and rehabilitation of cells back into the replicative pool, it will enact a programme of cell cycle arrest and repair. p53 is also able to upregulate antioxidant expression to combat accumulation of potentially damaging reactive oxygen species, such as SESN1, SESN2 and TIGAR [60]. As p53 continues to be extensively studied, additional transcriptional targets continue to be uncovered, including those relating to potential physiological roles, such as fertility (LIF) [61], pigmentation(KITLG) [57], differentiation and ageing [50].

It is worth mentioning that the majority of tumour-derived p53 mutations occur within the DNA binding domain, and therefore impact upon p53's transcription

factor function. The most commonly altered p53 residues in cancer, termed 'hotspot mutations' are R175, G245, R248, R249, R273 and R282. These mutations can impact p53's contact with its response elements (termed contact mutants) or disrupt proper DBD folding (termed structural mutants) [62]. As well as being able to disrupt DNA binding, these mutations can confer gain-of-function capabilities, which have been shown to play important roles in cancer, disease progression, invasiveness and metastasis [63].

1.4.4 Additional functions of p53

In addition to its function as a transcription factor, p53 has been attributed a range of other functions. p53 has been directly implicated in apoptosis, through interaction with members of the BCL-2 family of apoptotic regulators at the mitochondria, followed by the induction of mitochondrial outer membrane permeabilisation [59]. p53 has also been linked to necrotic cell death, again through its direct interaction with mitochondrial components [59]. Interestingly, cytoplasmic p53 has been implicated in the repression of autophagy, in stark contrast to its upregulation of pro-autophagy genes [59].

Despite being subject to intensive study over the past 40 years, emerging roles of p53 are still being discovered and characterised, providing additional insights into the tumour suppressor and physiological functions of p53. One such function is the observed ability of p53 to bind RNA.

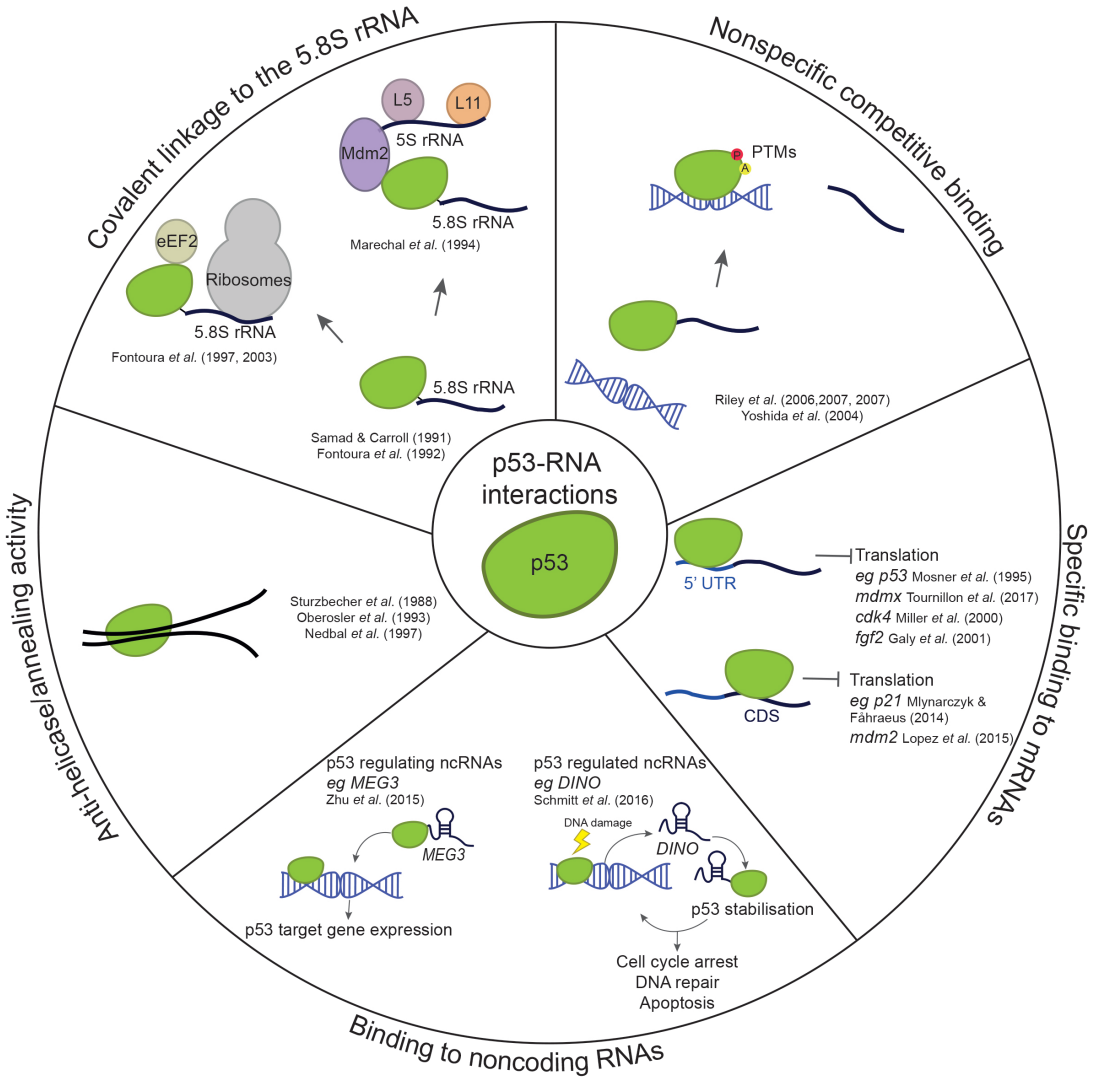


Figure 1.5: Summary of reported p53-RNA interactions Schematic representation of p53-RNA interactions reported to date (clockwise). The first p53-RNA interaction to be identified was an unusual covalent linkage with the 5.8S rRNA in SV40 transformed cells. Non-specific binding to RNA that displaces binding of single and double stranded DNA species was demonstrated, along with specific binding and translational repression of several mRNA species. Recently, ncRNAs have been shown to bind to and modulate p53 activity, including via autoregulatory loops. Much of the proposed impact of p53 binding on RNA is proposed to be due to p53's ability to enhance RNA annealing, with p53 termed an RNA chaperone.

1.5 p53-RNA interactions

p53-RNA interactions have been investigated for over 30 years, with the major findings to date summarised in Figure 1.5.

The first suggestion of a p53-RNA interaction was a covalent linkage between

the C-terminus of p53 and ribonucleoside monophosphates in SV40 transformed cells [64]. The covalently bound RNA species was subsequently identified as the 5.8S rRNA, linked to p53's Ser289 residue [65, 66]. This interaction was further expanded by Fontoura *et al.* [67], who coimmunoprecipitated this p53-5.8S RNA species with a subset of ribosomes, along with *p53* mRNA. Further association of p53 with the eukaryotic elongation factor 2 (eEF2) in a subset of ribosomes compounded the association between p53 and the translational machinery, suggesting a role in translation elongation [68]. A large amount of this work was performed in transformed cell lines, including lines expressing the SV40 large T antigen, which is known to repress the p53 pathway [69]. Subsequent works have been unable to reproduce this covalent linkage; suggesting a possible cell type or transformation-specific interaction.

In addition, Marechal *et al.* [70] identified an RNP complex comprising the 5S rRNA, L5 ribosomal protein, MDM2 and p53 (5S RNA-L5-mdm-2-p53 RNP). Marechal *et al.* suggest the possibility of the presence of the 5.8S rRNA within this RNP, due mainly to the reports of covalent interaction between p53 and the 5.8S rRNA, combined with the propensity of the L5 ribosomal protein and 5S rRNA to form a ternary complex with the 5.8S rRNA in solution [70]. The presence of this larger RNP remains largely speculative, with little subsequent experimental evidence of its existence.

1.5.1 Anti-helicase & annealing activity

p53's first described RNA binding function was its anti-helicase function, via observed inhibition of the SV40 large T-antigen [71], and other RNA and DNA heli-

cases [72]. Oberosler *et al.* [72] attributed this activity to recombinant tetrameric p53's ability to promote annealing of single stranded (ss) RNAs and DNAs. p53 was able to induce conformational changes in the RNA, which alluded to a role as an RNA chaperone; able to facilitate RNA folding and the stabilisation of certain RNA secondary structures. Interestingly, an amino acid within the core DNA binding domain was suggested as key for this RNA binding and annealing activity. Subsequent work by Nedbal *et al.* [73] demonstrated between 490-1600 fold increase in RNA annealing in the presence of p53. This was attributed to sequence independent RNA binding via the basic domain of the C-terminus, which demonstrated very little impact on RNA secondary structure; in contrast to the earlier findings of Oberosler *et al.*

1.5.2 Non-specific versus specific interactions

As investigations into the function of p53's RNA binding activities intensified, the field became split between two extremes, debating the specificity of the p53-RNA interaction.

On the one hand, Yoshida *et al.* [74], proposed p53's RNA binding activity to be a sequence and structurally non-specific mechanism by which inactive cytoplasmic p53 could be maintained in a latent state. Phosphorylation of Ser392, which is frequently associated with p53 activation, was shown to reduce RNA binding, and enhance p53 oligomerisation; restoring sequence specific DNA binding at the core DNA binding domain. Riley *et al.* [75] identified p53 as able to bind RNA in a yeast three-hybrid system, but observed no sequence or structural specificity. Additional *in vitro* work with recombinant p53 confirmed the CTD was necessary and sufficient

for p53 binding, which was disrupted upon PTM (acetylation of 4 lysine residues in this case), and that such binding could inhibit sequence specific DNA binding at the core domain [76].

Conversely, a group of investigators began revealing the supposedly specific nature of the p53-RNA interaction. Galy *et al.* [77], Miller *et al.* [78], Tournillon *et al.* [79] and Mosner *et al.* [80], proposed the translational repression of *FGF-2, Cdk4, MDMX* and *p53*, respectively, through binding of p53 to regulatory elements within transcript 5'-UTRs. Mechanistically, they proposed p53's anti-helicase/RNA annealing activity may induce conformational changes within the 5'-UTRs that inhibits translation. On the other hand, Lopez *et al* [81], along with Mlynarczyk and Fahraeus [82], suggest an alternative mechanism, involving structure-mediated binding of p53 to transcript coding sequences in order to repress *p21* and *MDM2* translation under endoplasmic reticulum (ER) stress.

Postulated functions of the interactions of these selected candidate RNAs and p53 involve fine-tuning of p53 responsive genes under both physiological and stress conditions, with the possibility of post-transcriptional feedback loops.

1.5.3 Non-coding RNAs and p53

Recent advances in the p53-RNA field have come from the 'non-coding genome', with multiple non-coding RNAs (ncRNAs) identified as potential interaction partners, including p53-induced ncRNAs. An example includes the lncRNA *DINO* (Damaged Induced Noncoding), which is able to bind p53 and amplify its stabilisation in response to DNA damage [83]. Such feedback loops have been observed with other ncRNAs, including miRNAs [84]. Additionally, ncRNAs have been shown to exert

their function on p53, including the tumour suppressive *MEG3* lncRNA (Maternally Expressed Gene 3). *MEG3* has been shown to directly interact with p53, reportedly through the core DNA binding domain of p53, which increases p53's stability, and subsequent transcriptional activation of targets such as SESN1 and TGFA [85, 86]. As the importance of ncRNAs, such as miRNAs and long non-coding (lncRNAs) in the p53 pathway becomes increasingly apparent, it is likely that more of these interactions will be identified.

1.5.4 p53 binding interface

Considerations of the p53 domains involved in RNA binding have lead to two main hypotheses; RNA binds to the core DNA binding domain of p53, or alternatively, to the C-terminus. Both domains have an intrinsic ability to bind nucleic acids, but whether this translates to functional RNA binding remains to be elucidated. As a predominantly regulatory domain, the C-terminus of p53 is subject to diverse post-translational modifications, as well as providing a platform for interaction with multiple nucleic acid and protein species, representing potential mechanisms by which RNA-binding could be regulated [87].

To date, technological limitations have prevented full characterisation of the p53-RNA interaction. Previous studies have relied on *in vitro* and acellular systems, with heterologously expressed or recombinant p53 proteins (predominantly bacterial or insect) bearing little resemblance to physiological p53 in terms of concentrations used and post-translational modification status [87]. Additionally, previous approaches have either addressed p53's overall ability to bind RNA without further validation of RNA targets, or focused on characterisation of a single p53-RNA interaction *in vitro*.

Consequently, studies to date have generally assessed the biochemical propensity of p53 to bind RNA. Still lacking therefore, is an understanding of whether p53 binds RNAs in living cells, and the subsequent specificity and function of such physiological interactions.

1.6 Aims

Despite being one of the most studied proteins, relatively little is known regarding p53's interactions with RNA in a physiological context. This project aims to address this lack of understanding, specifically by:

- Identifying p53-RNA interactions in living cells using novel techniques.
- Characterising the p53-RNA interaction interface, including the p53 RNA interactome (RNA-centric approaches) and p53 RNA binding domains (protein-centric approaches).
- Evaluating the importance of RNA in p53's protein interaction network.

This work aims to provide valuable insight into the underlying mechanisms of p53-RNA interactions, along with potential functional implications.

2

Materials

2.1 Materials

2.1.1 Chemicals, reagents, and consumables

Table 2.1: Reagents used in this study.

Reagent/Consumable	**Supplier**	**Code**
1 cm glass pre-filters	Whatman	1823010
2log DNA ladder	NEB	N3200S
4.4.5.5-D4-L-Lysine	Silantes GmbH	211104113
6% TBE precast gels 10 wells	Invitrogen	EC6265BOX
6% TBE precast gels 15 wells	Invitrogen	EC62655BOX
6% TBE-Urea precast gels 10 wells	Invitrogen	EC6865BOX
6x purple gel loading dye	NEB	B7024S
6x purple gel loading dye no SDS	NEB	B7025S
96-well microplate with flat μClear bottom	Greiner Bio-One	655986
Accuprime Supermix 1 enzyme	Invitrogen	12342-010
Acetic acid glacial	VWR Chemicals	20104.334
Acid-Phenol:Chloroform pH 4.5 (with IAA, 125:24:1)	Invitrogen	AM9722
Acylamide:Bisacrylamide solution	Serva	10688.01
AEBSF Hydrochloride	BioChemica	A1421,0100
Agarose	MP Biomedicals	AGAP0100
Agilent High Sensitivity DNA Reagents	Agilent Technologies	5067-4627

Continuation of Table 2.1

Reagent	**Supplier**	**Identifier**
Agilent RNA 6000 Pico reagents	Agilent Technologies	5067-1514
Amicon Ultra 0.5 ml centrifugal filters	Millipore	UFC510024
Ammonium acetate	Honeywell	9688
Ammonium persulfate	Acros Organics	201531000
Antarctic phosphatase	NEB	M0289S
BamHI-HF	NEB	R3136S
Bensonase	EMD Millipore	70746
Blasticidin S	TOKU-E	B001
Bovine serum albumin (BSA)	Sigma-Aldrich	A7030
Cell scrapers	Fischer brand	08-100-241
CircLigase II ssDNA Ligase	Epicentre – Biozym	CL9021K
Clarity Western ECL Blotting Substrate	Bio-Rad	1705061
Colour prestained protein standard, broad range	NEB	P7719
Costar Spin-X centrifuge tube filters	Corning	CLS8160
Dialysed fetal bovine serum	Silantes GmbH	281000900
DMEM lacking phenol-red	Gibco	21063-029
DMSO	Sigma-Aldrich	D2260
DNA Lobind tubes 1.5 ml	Eppendorf	0030108051
Doxycycline	Sigma-Aldrich	D9891

Continuation of Table 2.1

Reagent	**Supplier**	**Identifier**
DL-dithiothreitol (DTT)	Sigma-Aldrich	D1532
Dulbecco's Modified Eagle Medium	Gibco	11995-065
Ethanol absolute	VWR chemicals	20821.33
Ethylenediaminetetraacetic acid (EDTA)	VWR chemicals	100935V
Etoposide	Sigma-Aldrich	E1383
FastDigest BamHI	Biozym	CL9021K
Fetal Bovine Serum (FBS)	Sigma-Aldrich	F9665
GFP-trap agarose beads	Chromotek GmbH	gta-10
Glass Microfiber filters	Whatman	1823-010
Glycerol	Sigma-Aldrich	G7757
Glycine	Sigma-Aldrich	G8898
Glycoblue	Invitrogen	AM9515
Hard-Shell PCR plates, 96 well	Bio-Rad	HSP9645
HindIII-HF	NEB	R3104S
Hygromycin B	TOKU-E	H011
IDC reagent	Thermo Fisher	22663
IGEPAL CA-630	Sigma-Aldrich	I3021
IRdye-800CW-DBCO	LI-COR	929-50000
IV 5 ml syringes luer lock	Medicina	IVL05
L-Arginine HCl 13C	Silantes GmbH	201204102
L-Arginine HCl 13C, 15N	Silantes GmbH	201604102

Continuation of Table 2.1

Reagent	**Supplier**	**Identifier**
L-Glutamine	Gibco	25030-081
L-Lysine HCl 13C, 15N	Silantes GmbH	211604102
Lithium chloride	Sigma-Aldrich	62476
Lithium-dodecyl sulfate	Sigma-Aldrich	L9781
Lobind 1.5 ml tubes	Sarstedt	72.706.600
Low Molecular Weight DNA ladder	NEB	N3233S
Magneisum acetate tetrahydate	Sigma-Aldrich	M5661
Magnesium chloride hexahydrate	Sigma-Aldrich	M2670
Max Efficiency DH5a Competent cells	Invitrogen	18258012
MaXtract High Density tubes	Qiagen	129056
Methanol	VWR Chemicals	20847.307
Microlance 27G needle	Becton Dickinson	2023-02
Midori green	Bulldog Bio	MG04
N,N,N',N'-Tetramethylethylenediamine	Sigma-Aldrich	T9281
NEBNext sample purification beads	NEB	E7767
NextSeq 500/550 High Output V2 kit, 75 cycles	Illumina	FC-404-2002
NextSeq 500/550 High Output V2.5 kit, 75 cycles	Illumina	20024906
NotI-HF	NEB	R3189L
Novex Hi-Density TBE Sample Buffer (5X)	Invitrogen	LC6678

Continuation of Table 2.1

Reagent	**Supplier**	**Identifier**
Novex NuPAGE 4-12% Bis-Tris gradient gels	Invitrogen	NP0322BOX
Novex TBE-Urea Sample Buffer (2X)	Invitrogen	LC6876
Nuclease free water	Thermo Fisher	AM9937
NuPAGE 4x LDS sample buffer	Invitrogen	NP0008
NuPAGE MOPS SDS running buffer (20x)	Invitrogen	NP0001
Nutlin	Sigma	SML0580
Odyssey Blocking Buffer (PBS)	LI-COR	1704159
Oligo(dT)25 beads	NEB	S1419
PBS 1X, pH 7.4	Gibco	10010023
Penicillin-Streptomycin	Sigma-Aldrich	P4458
Phusion High Fidelity DNA polymerase	NEB	MO530S
Pierce 660 nm Protein Assay Kit	Thermo Fisher	22662
Pierce control agarose resin	Thermo Fisher	26150
Poly(ethylene glycol) 400 (PEG400)	Sigma-Aldrich	202398
Precision plus protein dual colour standard	Bio-Rad	1610374
Proteinase K	Roche	3115887001
Qubit Protein Assay Kit	Invitrogen	Q33211
Restore western blot stripping buffer	Thermo Fisher	21059
Ribolock Rnase inhibitor	Thermo Fisher	EO0381
RNase A	Sigma-Aldrich	R4642
RNase I	Invitrogen	AM2295
RNase T1	Sigma-Aldrich	R1003

Continuation of Table 2.1

Reagent	**Supplier**	**Identifier**
RPMI 1640	Lonza Bioscience	BE12-702F
SILAC RPMI media	Silantes GmbH	283001300
Sodium acetate 3 M	Alfa Aesar	J61928
Sodium chloride	Sigma-Aldrich	31434
Sodium deoxycholate	Sigma-Aldrich	D6750
Sodium dodecyl sulfate	Sigma-Aldrich	L3771
Sodium hydroxide 1 M	Honeywell	011-002-00-6
Superscript III	Invitrogen	18080-044
Sybrgold	Invitrogen	S11494
T4 DNA ligase	NEB	M0202S
T4 Polynucleotide Kinase	NEB	M0201S
T4 RNA Ligase I	NEB	M0204S
Trans-Blot Turbo Midi Nitrocellulose Transfer Packs	Bio-Rad	1704159
Trans-Blot Turbo RTA Midi PVDF Transfer Kit	Bio-Rad	1704273
Triton X-100	MP Biomedicals	4807423
Trizma base	Sigma-Aldrich	93352
Trypan Blue Stain (0.4%)	Invitrogen	T10282
Trypsin EDTA	Sigma-Aldrich	T3924
TURBO DNase	Invitrogen	AM2238

Continuation of Table 2.1

Reagent	**Supplier**	**Identifier**
Tween-20	Sigma-Aldrich	P1379
Urea	Sigma-Aldrich	U0631
X-tremeGENE 9 DNA transfection reagent	Sigma-Aldrich	6365787001
XhoI	NEB	R0146L
Zeocin	Invitrogen	R25001

2.1.2 Buffers and solutions

Table 2.2: Buffers and solutions.

Buffer	**Composition**
4x ligation buffer	200 mM Tris HCl pH 7.5, 40 mM $MgCl_2$, 40 mM DTT
5x PNK buffer	350 mM Tris HCl pH 6.5, 50 mM $MgCl_2$, 25 mM DTT
High salt buffer	500 mM NaCl, 1 mM $MgCl_2$, 20 mM Tris HCl pH 7.5, 0.05% NP40, 0.1% SDS, 1 mM DTT, 1x AEBSF
High salt IP buffer	500 mM NaCl, 10 mM Tris HCl pH 7.5, 5 mM $MgCl_2$, 1% Triton X-100, 1x AEBSF, 1 mM DTT

Continuation of Table 2.2

Buffer	**Composition**
iCLIP GFP lysis buffer	100 mM NaCl, 5 mM $MgCl_2$, 10 mM Tris HCl pH 7.5, 0.5% NP40, 0.1% SDS, 0.5% Na-deoxycholate, 1 mM DTT, 1x AEBSF
Interactome capture buffer 1	500 mM LiCl, 20 mM Tris HCl pH 7.5, 0.1% LiDS, 1 mM EDTA, 0.1% NP40, 5 mM DTT
Interactome capture buffer 2	500 mM LiCl, 20 mM Tris HCl pH 7.5, 1 mM EDTA, 0.01% NP40, 5 mM DTT
Interactome capture buffer 3	200 mM LiCl, 20 mM Tris HCl pH 7.5, 1 mM EDTA, 5 mM DTT
Interactome capture elution buffer	20 mM Tris HCl pH 7.5, 1 mM EDTA
Interactome capture lysis buffer	500 mM LiCl, 20 mM Tris HCl pH 7.5, 0.5% LiDS, 1 mM EDTA, 0.1% NP40
Low salt IP buffer	150 mM NaCl, 10 mM Tris HCl pH 7.5, 5 mM $MgCl_2$, 1% Triton X-100, 1x AEBSF, 1 mM DTT
Lysogeny broth (LB)	For 1 L: 10 g Bacto tryptone (Difco), 5 g Bacto yeast extract, 5 g NaCl
Medium salt buffer	250 mM NaCl, 20 mM Tris HCl pH 7.5, 1 mM $MgCl_2$, 0.05% NP40, 1x AEBSF
PBS-T	1X PBS + 0.1% Tween-20

Continuation of Table 2.2

Buffer	**Composition**
PK buffer	50 mM NaCl, 100 mM Tris HCl pH 7.5, 10 mM EDTA
PK-urea buffer	50 mM NaCl , 100 mM Tris HCl pH 7.5, 10 mM EDTA, 7M Urea
PNK wash buffer	20 mM Tris HCl pH 7.5, 10 mM $MgCl_2$. 0.2% Tween-20
RIPA buffer	150 mM NaCl , 50 mM Tris HCl pH 7.5, 0.1% SDS, 0.5% Na-deoxycholate, 1% Triton X-100 , 1x AEBSF
SDS running buffer	192 mM Glycine, 25 mM Tris base, 0.1 % SDS
TAE buffer	40 mM Tris base, 1 mM EDTA-Na, 20 mM acetic acid, pH 8.5
TE buffer	10 mM TrisHCl, 1 mM EDTA in diethylpyrocarbonate (DEPC) treated water

2.1.3 Commercial kits

Table 2.3: Commercial kits. .

Kit	**Supplier**	**Identifier**
5' DNA adenylation kit	NEB	E2610S
Luna One-Step RT-qPCR kit	NEB	E3005S

Continuation of Table 2.3

Kit	**Supplier**	**Identifier**
NEBNext Multiplex Oligos for Illumina (Index Primers Set 1)	NEB	E7335
NEBNext Multiplex Oligos for Illumina (Index Primers Set 2)	NEB	E7500S
NEBNext ultra II directional library prep kit	NEB	E7760S
QIAGEN Plasmid Maxi kit	Qiagen	12162
QIAprep spin Miniprep kit	Qiagen	27104
QIAquick Gel Extraction kit	Qiagen	28704
QIAquick Nucleotide Removal kit	Qiagen	28304
QIAquick PCR Purification kit	Qiagen	28104
RNase free DNase set	Qiagen	79254
RNeasy mini kit	Qiagen	74104
SilverQuest Silver Staining Kit	Thermo Fisher	LC6070

2.1.4 Antibodies

Table 2.4: Antibodies.

Antibody	**Host species**	**Supplier/Code**
Anti-p53 (DO-1)	Mouse	Santa Cruz/sc-126
Anti-GFP (3H9)	Rat	Chromotek/3h9-20
Anti-β actin	Mouse	Sigma/A2228

Continuation of Table 2.4

Antibody	**Host species**	**Supplier/Code**
Anti-PTBP1	Mouse	Sigma
Anti-mouse IgG Alexa Fluor 488	Donkey	ThermoFisher Scientific A-21202
anti-mouse IgG, HRP	Goat	ThermoFischer Scientific G-21040
Anti-rat IgG, Alexa Fluor 488	Goat	ThermoFisher Scientific A-11006

2.1.5 Plasmids

Table 2.5: Plasmids. .

Plasmid	**Backbone**	**Source**
pCDNA5/FRT/TO	pcDNA5/FRT/TO	ThermoFisher Scientific
pOG44 Flp-Recombinase Expression vector	pOG44	ThermoFisher Scientific
p53 cDNA	pcDNA3	Lu lab
eGFP-p53	pcDNA5/FRT/TO	Castello lab
p53-eGFP	pcDNA5/FRT/TO	Castello lab
FLAG-HA-p53	pcDNA5/FRT/TO	Castello lab
p53-FLAG-HA	pcDNA5/FRT/TO	Castello lab

2.1.6 Oligonucleotides and primers

Oligonucleotides used for cloning

Table 2.6: Oligos used for cloning. .

Name	**Sequence**
N-terminal FLAG-HA tag	F-CCCAAGCTTGAGGAGCCGCAGTCAGATCCT R-CCGCTCGAGTCAGTCTGAGTCAGGCCCTTC
N-terminal eGFP tag	F-CGCGGATCCGAGGAGCCGCAGTCA R-TTTATAGCGGCCGCTCAGTCTGAGTCAGGCCC
C-terminal tags (eGFP and FLAG-HA)	F-CCCAAGCTTATGGAGGAGCCGCAGTCAGAT R-CCGCTCGAGGTCTGAGTCAGGCCCTTCTGT

Oligonucleotides for RT-qPCR

Table 2.7: Oligos used for RT-qPCR.

Name	**Sequence**
GAPDH	F-AGCCACATCGCTCAGACAC R-GCCCAATACGACCAAATCC
p21	F-TGAGCCGCGACTGTGATG R-GTCTCGGTGACAAAGTCGAAGTT
p53	F-CCCCAGCCAAAGAAGAAAC R-AACATCTCGAAGCGCTCAC
PUMA	F-GACCTCAACGCACAGTACGA R-TGGGTAAGGGCAGGAGTC

iCLIP oligonucleotides and primers

Table 2.8: Oligos and primers used during the iCLIP protocol.

Name	Sequence
L3 linker	/5Phos/AGATCGGAAGAGCGGTTCAGAAAAAAA-PAAAAA/iAzideN/AAAAAAAAAAAA/3Bio/
Cut oligo	GTTCAGGATCCACGACGCTCTTCAAAA
P5 Solexa primer	AATGATACGGCGACCACCGAGATCTACACTCTT-TCCCTACACGACGCTCTTCCGATCT
P3 Solexa primer	CAAGCAGAAGACGGCATACGAGATCGGTCTCGG-CATTCCTGCTGAACCGCTCTTCCGATCT

Table 2.9: iCLIP Rclip primers.

Name	Sequence
#1	/5Phos/NNAACCNNNAGATCGGAAGAGCGTCGTGGATCCTGAACCGC
#2	/5Phos/NNACAANNNAGATCGGAAGAGCGTCGTGGATCCTGAACCGC
#3	/5Phos/NNATTGNNNAGATCGGAAGAGCGTCGTGGATCCTGAACCGC
#4	/5Phos/NNAGGTNNNAGATCGGAAGAGCGTCGTGGATCCTGAACCGC
#5	/5Phos/NNCGCCNNNAGATCGGAAGAGCGTCGTGGATCCTGAACCGC
#6	/5Phos/NNCCGGNNNAGATCGGAAGAGCGTCGTGGATCCTGAACCGC
#7	/5Phos/NNCTAANNNAGATCGGAAGAGCGTCGTGGATCCTGAACCGC
#8	/5Phos/NNCATTNNNAGATCGGAAGAGCGTCGTGGATCCTGAACCGC
#9	/5Phos/NNGCCANNNAGATCGGAAGAGCGTCGTGGATCCTGAACCGC
#10	/5Phos/NNGACCNNNAGATCGGAAGAGCGTCGTGGATCCTGAACCGC
#11	/5Phos/NNGGTTNNNAGATCGGAAGAGCGTCGTGGATCCTGAACCGC
#12	/5Phos/NNGTGGNNNAGATCGGAAGAGCGTCGTGGATCCTGAACCGC
#13	/5Phos/NNTCCGNNNAGATCGGAAGAGCGTCGTGGATCCTGAACCGC
#14	/5Phos/NNTGCCNNNAGATCGGAAGAGCGTCGTGGATCCTGAACCGC
#15	/5Phos/NNTATTNNNAGATCGGAAGAGCGTCGTGGATCCTGAACCGC

Continuation of Table 2.9

Name	Sequence
#16	/5Phos/NNTTAANNNAGATCGGAAGAGCGTCGTGGATCCTGAACCGC

2.1.7 Cell lines

Parental cell lines

Table 2.10: Parental cell lines.

Cell line	Cell type	Source
Hek293 Flp-In T-REx	Human embryonic kidney cells	Castello lab
MCF-7	Human breast adenocarcinoma	Bond lab

Genetically modified cell lines

Table 2.11: Self-generated modified cell lines.

Cell line	Cell type
Hek293 Flp-In T-REx GFP	Hek293 Flp-In T-REx
Hek293 Flp-In T-REx hnRNPC-GFP	Hek293 Flp-In T-REx
Hek293 Flp-In T-REx GFP-p53	Hek293 Flp-In T-REx

3

Methods

3.1 Cell biology

3.1.1 Cell culture

All cells were cultured under sterile conditions at 37^{o}C with 5% CO_2. MCF-7 cells were obtained as part of the NCI60 cell panel, a panel of 59 human cancer cell lines, from the NC-Frederick cancer DCTC tumour\cell line repository, Maryland. MCF-7s were cultured in Roswell Park Memorial Institute (RPMI) 1640 medium supplemented with L-Glutamine, 10% Foetal Bovine Serum (FBS) and 1x Penicillin/Streptomycin (P/S). HEK293 Flp-In T-REx and HeLa Flp-In T-REx lines were cultured in Dulbecco's Modified Eagle Medium (DMEM) with 10% FBS, 1x P/S, 15 μg/ml Blasticidin S and 100 μg/ml Zeocin. Cells were passaged as required to maintain density below 90-95%, by removing media, washing with phosphate-buffered saline (PBS), and adding the relevant volume of Trypsin/EDTA solution (0.5 ml T25 flask, 1 ml T75 flask, 2 ml T175 flask). Cells were incubated in trypsin solution for up to 10 minutes in a 37^{o}C incubator, before being resuspended in the relevant medium, and either plated at the required density for experimental use, or replated. When required, cells were counted using Trypan Blue Stain (0.4%) and the Countess II FL automated cell counter (both Invitrogen).

3.1.2 Cell culture in SILAC media

For quantitative proteomic experiments, MCF-7 cells were cultured in SILAC RPMI media, supplemented with 10% dialysed FBS and 1x P/S, along with isotopic labelled lysine and arginine (Silantes GmbH amino acids: L-Arginine 13C,15N labeled; L-Arginine 13C labeled; L-Lysine 13C,15N labeled; 4.4.5.5.-D4-L-Lysine). Label in-

corporation was confirmed by mass spectrometry of whole cell lysates, ensuring >98% incorporation of the isotopic labelled amino acids.

3.1.3 Stable transfection of Flp-In T-REx cells

Hek293 Flp-In T-REx cells were seeded in 6 well plates to allow 80% confluency at the time of transfection. Transfection reagent mix was prepared by adding X-tremeGENE 9 DNA transfection reagent to optiMEM in a 3:1 ratio, before mixing. A total of 1 μg of pOG44 Flp-recombinase expression vector and the relevant insert-containing pcDNA5 FRT/TO plasmid, were then added to the transfection reagent mix, which was mixed and incubated at room temperature for 15 minutes. The transfection complex was then added dropwise to the cells, ensuring even distribution over the plate by gentle mixing. Cells were then incubated for 24 hours, before passage into a T25 flask, accompanied by the addition of selective media (replacing Zeocin with 150 μg/ml Hygromycin B). Induction of the relevant eGFP fusion protein was then achieved by supplementing media with 1 μg/ml Doxycycline.

3.1.4 Platereader assay

Cells were seeded at a density of 8x10^4 cells per well in a 96-well microplate with flat μClear bottom (Greiner Bio-One) in DMEM lacking phenol-red, supplemented with 10% FBS, 1x P/S, 15 μg/ml Blasticidin S and either Zeocin or 150 μg/ml Hygromycin B, depending on the cell type. Once cells had reached approximately 60% confluency, the relevant induction or control treatment was added, and cells incubated at 37^{o}C with 5% CO_2 in a CLARIOstar fluorescence plate reader (BMG Labtech). eGFP signal was monitored over a 24 hour period by measuring fluores-

cence (eGFP: excitation 470 nm, emission 515 nm) every 15 minutes.

3.2 Biochemical and molecular biology techniques

3.2.1 Cloning

To generate tagged p53 constructs, the p53 cloning sequence was amplified from a p53/pcDNA3 vector (kind gift from the Lu lab), using specific PCR primers including specified restriction sites (see Table 3.1) according to the Phusion DNA High-Fidelity polymerase (NEB) protocol, as follows.

Table 3.1: Reaction mix for cloning PCR.

Component	Volume
Template DNA	1 μl
5x Phusion HF reaction buffer	10 μl
dNTP mix (10 mM)	1 μl
Forward primer (10 μM)	2.5 μl
Reverse primer (10 μM)	2.5 μl
Phusion DNA polymerase	0.5 μl
Nuclease free water	32.5 μl

Table 3.2: Cloning PCR cycling conditions.

Step	Temperature	Time (mm:ss)	Cycles
Initial denaturation	98^{o}C	00:30	1
Denaturation	98^{o}C	00:10	
Melting	66^{o}C	00:20	25
Extension	72^{o}C	01:30	
Final extension	72^{o}C	05:00	1

PCR products were then combined with 6x purple gel loading dye with SDS (NEB), and run on a 1% agarose gel (1% agarose (MP Biomedicals) dissolved in 1x TAE buffer) with 1:20000 midori green DNA stain (Bulldog bio), along with 5 μl 2-log DNA ladder (NEB). Gels were run in 1x TAE buffer, at 150 V, before visualisation with either a UV transilluminator, or using the LI-COR Odyssey Fc imaging

system, using the 600 nm channel. The resulting bands were excised and purified using the QIAquick gel extraction kit (Qiagen), according to the manufacturer's protocol, with the final elution performed using 30 μl nuclease free water.

Resulting p53 inserts were then digested using the following restriction enzymes.

Table 3.3: Restriction enzymes used for p53 insert digestion.

Insert	Restriction enzymes
N-terminal FLAG-HA p53	HindIII,XhoI
N-terminal eGFP p53	BamHI, NotI
C-terminal FLAG-HA p53	HindIII, XhoI
C-terminal eGFP p53	HindIII, XhoI

The pcDNA5 FRT/TO plasmid was also digested accordingly. The following restriction digests were set up, and incubated for 1 hour at 37^{o}C.

Table 3.4: Insert restriction digest.

Component	Volume
Insert DNA	15 μl
10x NEB buffer 2	4 μl
Restriction enzyme 1	1 μl
Restriction enzyme 2	1 μl
Nuclease free water	9 μl

Table 3.5: Plasmid restriction digest.

Component	Volume
pcDNA5 FRT/TO plasmid DNA	4 μl
10x NEB buffer 2	4 μl
Restriction enzyme 1	1 μl
Restriction enzyme 2	1 μl
Nuclease free water	30 μl

1 μl Antarctic phosphatase (NEB) was then added to the reactions, along with an additional 4 μl 10x NEB buffer 2. Reactions were incubated for a further hour at 37^{o}C, followed by heat inactivation at 70^{o}C for 5 minutes.

Samples were run on a 1% agarose gel and gel extracted as previously, and quantified using the nanodrop spectrophotometer (Thermo Fisher).

Ligation reactions were then set up using a 1:3 plasmid:insert molar ratio. The reaction mixtures were set up (see Table 3.6), and incubated at 22^{o}C for 1 hour.

Table 3.6: Ligation reaction mixture.

Component	Volume
Plasmid DNA	X μl
Insert DNA	Y μl
10x T4 buffer	2 μl
T4 DNA ligase	1 μl
Nuclease free water	Up to 20 μl

Resulting DNA constructs were sent for sequencing (Source Bioscience) to confirm the presence and correct orientation of the tagged p53 within the pcDNA5 FRT-TO plasmid.

3.2.2 Plasmid preperation

Plasmids whose sequence had been validated were transformed into MAX Efficiency DH5α Competent Cells (Invitrogen), according to the manufacturer's protocol. Briefly, competent cells were incubated on ice with the plasmid of interest for 30 minutes, prior to heat shock at 42^{o}C for 45 seconds. Cells were returned to ice for 2 minutes, before the addition of 900 μl S.O.C medium and shaking at 37^{o}C for 1 hour (225 rpm). Resulting reactions were diluted to varying degrees and 100 μl spread onto selective LB agar containing 50 μg/ml of ampicillin. These were incubated overnight at 37^{o}C.

The next day, 3-5 colonies were selected from each plate, and transferred into 7 ml of LB containing 50 μg/ml of ampicillin, which were incubated overnight at 37^{o}C with shaking at 225 rpm.

1 ml of starter culture was then taken and purified using the QIAprep spin miniprep kit (Qiagen) according to the manufacturer's protocol. Resulting DNA was

eluted in 30 μl nuclease free water, and quantified using the nanodrop. Constructs were then sent for final sequencing validation to confirm successful integration of the p53 insert into the pcDNA5 FRT/TO plasmid before transfection into cells.

3.2.3 Western Blotting

Samples were prepared by mixing with NuPAGE 4x LDS sample buffer (Invitrogen) and heated to 70^{o}C for 10 minutes. Proteins were then separated by SDS-PAGE on either precast Novex NuPAGE 4-12% Bis-Tris gradient gels (Invitrogen) or a SDS-polyacrylamide gel of the required percentage (8-10%). Proteins were then transferred to nitrocellulose or PVDF membranes using the Trans-Blot Turbo Transfer System (BioRad). Membranes were blocked in either 50% Odyssey Blocking Buffer in 0.1% PBS-T, or 5% BSA (depending on the antibody used) for 1 hour at RT. The membrane was incubated with primary antibody overnight at 4^{o}C. After washing in 0.1% PBS-T, membranes were incubated with the relevant secondary antibodies for 2 hours at RT. Membranes were then either washed in 0.1% PBS-T, or developed using ECL reagent according to the manufacturer's protocol, depending on the secondary antibody used. Membranes were then visualised using the LI-COR Odyssey Fc imaging system.

3.2.4 Silver staining

Samples were prepared by mixing with NuPAGE 4x LDS sample buffer (Invitrogen) and heated to 70^{o}C for 10 minutes. Proteins were then separated by SDS-PAGE on either precast Novex NuPAGE 4-12% Bis-Tris gradient gels (Invitrogen) or a SDS-polyacrylamide gel of the required percentage (8-10%). Gels were then rinsed

briefly in ultrapure water, before using the SilverQuest silver staining kit (Thermo Fisher), according to the manufacturer's quick staining protocol.

3.2.5 RNA isolation, cDNA synthesis and RT-qPCR

0.3 x 10^6 cells were seeded in each well of a 6-well plate, and once 80% confluent, were treated with the appropriate stimuli. Total RNA was then obtained using the RNeasy Mini kit with on column DNase digestion using the RNase-free DNase set (both Qiagen), according to the manufacturer's protocol. RT-qPCR was then carried out using the Luna Universal One-Step RT-qPCR kit (NEB) according to manufacturer's instructions, using gene-specific primers (Table 2.7) in a 96 well PCR plate (Bio-rad), using the BioRad CFX96 Real-Time system. The following PCR cycling conditions were used. Results were analysed using the REST relative expression software tool [88].

Table 3.7: Cycling conditions for RT-qPCR.

Step	**Temperature**	**Time (mm:ss)**	**Cycles**
Reverse transcription	55^{o}C	10:00	1
Initial denaturation	95^{o}C	01:00	1
Denaturation	95^{o}C	00:10	
Extension	60^{o}C	00:30	40-45
Melt curve	60 - 95^{o}C		1

3.3 RNase coupled proteomic analysis

3.3.1 RNase coupled immunoprecipitation

GFP and eGFP-p53 Hek293 Flp-In T-REx cells were seeded in 10 cm plates, in triplicate, and when approximately 80% confluent, eGFP or eGFP-p53 induced by addition of 1 μg/ml Doxycycline, or 1 μg/ml Doxycycline plus 10 μM nutlin or

etoposide, respectively. Culture media was removed, and cells washed in PBS. Cells were then collected in 1 ml PBS + AEBSF protease inhibitor, and pelleted by centrifugation (400 g, 5 minutes, 4^oC). Pellets were washed twice with PBS + AEBSF, before being stored at -80^oC. Pellets were then thawed on ice and lysed in 1 ml low salt IP buffer, before collecting the supernatant by centrifugation (16,000 g, 5 minutes, 4^oC). Inputs were taken. 25 μl agarose control bead slurry (Pierce) were equilibrated in low salt IP buffer, and added to the lysates with gentle rotation for 1 hour at 4^oC. The resulting precleared lysates were taken after centrifugation (2500 g, 2 minutes, 4^oC). 25 μl GFP-Trap agarose bead slurry (ChromoTek GmbH) were equilibrated in 1 ml low salt IP buffer before being added to the precleared lysates, and mixed with gentle rotation for 2 hours at 4^oC. The GFP-trap beads were washed once with low salt IP buffer, with the beads collected by centrifugation after each wash step (2500 g, 2 minutes, 4^oC). During the second wash step, samples designated +RNase had their low salt IP buffer supplemented with RNase A and RNase T1 (Sigma). Beads were incubated for 5 minutes at 37^oC with gentle agitation. Beads were subject to 3 additional washes with low salt IP buffer. Proteins were eluted from the beads by pH elution, consisting of resuspending the beads in 50 μl 0.2 M glycine for 45 seconds, followed by collection of the beads. The resulting supernatant was transferred to a new tube and neutralised by the addition of 5 μl 1 M Tris base pH 10.4. The elution step was performed twice for each sample, and the elutions combined.

3.3.2 Mass spectrometry

SP3, LC-MS analysis, protein identification and initial analyses were performed by Wael Kamel.

SP3

IP samples were processed using modified SP3 protocol ([89] Sielaff *et al* 2017) for mass spectrometric analysis. Beads were washed and reconstituted in water (10 μg solids / μl), and the bead mix stored at 4^{o}C, SP3 was performed in LoBind tubes (1.5 ml, Eppendorff). Samples were denatured (8 M urea, 100 mM TEAB pH 8), before reduction and nuclease digestion via the addition of 10 mM tris(2-carboxyethyl)phosphine (TCEP) , and 25 U Benzonase nuclease (Millipore). Samples were then alkylated in 50 mM iodacetamide. 10 μl bead mix was added to the lysate, before addition of acetonitrile (ACN) (80% final concentration). Samples were incubated at RT for 30 minutes. Beads were subsequently immobilised using a magnetic rack (2 minutes), before the supernatant was discarded. Two bead washes were performed with 1 ml 70% EtOH, followed by a 100% ACN rinse. These 3 wash steps were repeated 10 times to ensure removal of residual detergent. 1 μG trypsin in 50 mM TEAB was added to the beads, which were then digested overnight (37^{o}C, 500 rpm).

Samples were mixed prior to bead recovery by incubation on magnetic racks. Supernatants were transferred to new tubes. Beads were then incubated in 2% DMSO for 10 minutes to ensure complete recovery of peptides from the beads. A second supernatant was then retrieved by immobilisation on a magnetic rack, and combined with the first supernatant. Samples were acidified with 5% formic acid

prior to LC-MS analysis.

LC-MS/MS

Acidified peptides were analysed on an Ultimate 3000 ultra-HPLC system (Thermo Fisher Scientific) and electrosprayed into a QExactive mass spectrometer (Thermo Fischer Scientific). Peptides were initially trapped on a C18 PepMap100 pre-column (300 μm inner diameter $\times$ 5 mm, 100Å, Thermo Fisher Scientific) in solvent A (0.1% formic acid). Peptides were then separated on an in-house packed analytical column (75 μm inner diameter $\times$ 50 cm packed with ReproSil-Pur 120 C18-AQ, 1.9 μm, 120 Å, Dr. Maisch GmbH) using a linear 15-35% (vol/vol) ACN gradient, over 60 minutes, and a 200 nl/min flow rate. Full-scan mass spectra were acquired in the Orbitrap (scan range 350-1500 m/z, resolution 70000, AGC target 3 $\times$ 106, maximum injection time 50 ms) in a data-dependent mode. Subsequently, the 10 most intense peaks were selected for higher-energy collisional dissociation fragmentation at 30% of normalised collision energy. Higher-energy collisional dissociation fragmentation spectra were acquired in the Orbitrap (resolution 17500, AGC target 5 $\times$ 104, maximum injection time 120 ms) with first fixed mass at 180 m/z.

Protein identification and interaction analysis

Raw data were processed by MaxQuant (version 1.5.0.35 [90]) for peptide protein identification, using the integrated Andromeda search engine [91]. The mass spectra were searched against a human database (UniProt, version 01/16 [92]), along with the eGFP protein sequence, and a list of common contaminants. The following search parameters were used: Trypsin was chosen with full tryptic specificity and maximal two missed cleavage sites. Fixed modifications were set to carbamidomethyl

(C) and the variable modifications to acetylation (protein N-term, i.e. only the amino terminus of the protein) and oxidation (M). First search peptide tolerance was set to 50. Match between runs was applied. Default parameters were used for all others. Search results were filtered using a false discovery rate (FDR) of 0.01.

3.4 RNA interactome capture

3.4.1 RNA interactome capture

RNA interactome capture was performed based on the previously described protocol (Castello *et al.*, 2012, Castello *et al.*, 2013), as follows. MCF-7 cells were cultured in 5 x 15 cm dishes per condition in 3 biological replicates. Once 80% confluent, cells were treated with either 10 μM nutlin or etoposide. Cells were then washed with PBS before irradiation with 150mJ/cm^2 UV light at 254 nm, and lysed with 5 ml interactome lysis bufffer. Lysates were homogenised by passing at high speed through a 5 ml syringe fitted with a 27 G needle. This process was repeated until lysates were fully homogenised, and inputs were taken for total proteome, RNA and western blot analysis. 1.5 ml pre-equilibrated oligo(dT)25 magentic beads (NEB) were added, and lysates incubated, with gentle rotation, for 1 hour at 4^{o}C. Beads were collected by application of samples to a magnet, and the resulting supernatants stored at 4^{o}C. Beads were then incubated in 10 ml interactome capture lysis buffer, with gentle rotation at 4^{o}C, for 5 minutes. Washes were repeated twice each with interactome capture buffer 1 and interactome capture buffer 2. Beads were then washed twice with interactome capture buffer 3 at room temperature. Beads were then resuspended in 900 μl elution buffer, and incubated at 55^{o}C for 3 minutes with

agitation. Beads were then recycled according to the manufacturer's protocol for a second round of capture. Briefly, beads were resuspended in 900 μl 0.1 M NaOH, incubating for 5 minutes at 55^{o}C with agitation, and washing 3 times in interactome lysis buffer. The supernatant stored at 4^{o}C was then used for a second round of capture and the protocol repeated. RNA concentrations were then taken for each eluate, and the eluates combined.

3.4.2 Conventional protein analysis

Samples were resolved on SDS-PAGE and analyzed by western blotting using specific antibodies, the Li-Cor Odyssey system for visualization and the Image Studio Lite software (Li-Cor) for quantification, or silver staining.

3.4.3 Comparative RNA interactome capture

Comparative RNA interactome capture (cRIC) based on the previously described RNA interactome capture protocol and Garcia-Moreno *et al.*, 2018, was performed, with the following alterations: MCF-7 cells, previously grown in media with isotopic labeled amino acids, were seeded in three sets of 4x 15 cm dishes, each set with a different SILAC label. To correct for isotope-dependent effects, we permutated the SILAC labels between the three conditions in the five biological replicates. Cells were treated, UV irradiated, lysed in 4 ml interactome capture lysis buffer, and homogenised as previously. Inputs were taken, and the protein content of each sample measured using a kit compatible with ionic detergents (Thermo Fisher). Equal amounts of each of the three lysates were mixed and the final volume adjusted to 9 ml. The remainder of the interactome capture protocol was performed as previously,

but with a total of 3 interactome capture rounds. Eluate RNA concentration was measured via the nanodrop UV spectrophotometer, and all 3 eluates combined if RNA concentration was above 50 ng/μl.

3.4.4 Mass spectrometry

SP3, LC-MS analysis, protein identification and initial analyses were performed by Svenja Hester.

SP3

cRIC samples were processed using modified SP3 protocol ([89] Sielaff *et al* 2017) for mass spec analysis. Beads were washed and reconstituted in water (10 μg solids /μl), and the bead mix stored at 4^{o}C, SP3 was performed in LoBind tubes (1.5 ml, Eppendorff). Samples were denatured (8 M urea, 100 mM TEAB pH 8), before reduction and nuclease digestion via the addition of 10 mM TCEP , and 25 U Benzonase nuclease (Millipore). Samples were then alkylated in 50 mM iodacetamide. 10 μl bead mix was added to the lysate, before addition of ACN (80% final concentration). Samples were incubated at RT for 30 minutes. Beads were subsequently immobilised using a magnetic rack (2 minutes), before the supernatant was discarded. Two bead washes were performed with 1 ml 70% EtOH, followed by a 100% ACN rinse. These 3 wash steps were repeated 10 times to ensure removal of residual detergent. 1 μG trypsin in 50 mM TEAB was added to the beads, which were then digested overnight (37^{o}C, 500 rpm).

Samples were mixed prior to bead recovery by incubation on magnetic racks. Supernatants were transferred to new tubes. Beads were then incubated in 2%

DMSO for 10 minutes to ensure complete recovery of peptides from the beads. A second supernatant was then retrieved by immobilisation on a magnetic rack, and combined with the first supernatant. Samples were acidified with 5% formic acid prior to LC-MS analysis.

LC-MS/MS

Acidified peptides from cRIC experiments were analysed using an Ultimate 3000 ultra-HPLC system (Thermo Fisher Scientific) and electrosprayed into a QExactive mass spectrometer (Thermo Fischer Scientific). Peptides were initially trapped on a C18 PepMap100 pre-column (300 μm inner diameter $\times$ 5 mm, 100 Å, Thermo Fisher Scientific) in solvent A (0.1% formic acid). Peptides were separated on an in-house packed analytical column (75 μm inner diameter $\times$ 50 cm packed with ReproSil-Pur 120 C18-AQ, 1.9 μm, 120 Å, Dr. Maisch GmbH) using a linear 15-35% (vol/vol) ACN gradient (2 hrs for interactome capture samples) and a flow rate of 200 nl/min. Full-scan mass spectra were acquired in the Orbitrap (scan range 350-1500 m/z, resolution 70000, AGC target 3$\times$106, maximum injection time 50 ms) in a data-dependent mode. Subsequently, the 10 most intense peaks were selected for higher-energy collisional dissociation fragmentation at 30% of normalised collision energy. Higher-energy collisional dissociation fragmentation spectra were also acquired in the Orbitrap (resolution 17500, AGC target 5$\times$104, maximum injection time 120 ms) with first fixed mass at 180 m/z.

Protein identification and interaction analysis

Raw data were processed by MaxQuant [90] (version 1.5.0.35) for peptide protein identification, using the integrated Andromeda search engine [91]. The mass spectra

were searched against a human database (UniProt, version 01/16 [92]), along with the eGFP protein sequence, and a list of common contaminants. The following search parameters were used: Trypsin was chosen with full tryptic specificity and maximal two missed cleavage sites. Fixed modifications were set to carbamidomethyl (C) and the variable modifications to acetylation (protein N-term, i.e. only the amino terminus of the protein) and oxidation (M). First search peptide tolerance was set to 50. Match between runs was applied. Default parameters were used for all others. Search results were filtered using a false discovery rate (FDR) of 0.01.

3.5 CLIP-seq

GFP and eGFP-p53 Hek293 Flp-In T-REx cells were cultured in 1x 15 cm dish per condition, until approximately 80% confluent, when target proteins were induced via the addition of 1 μg/ml Doxycycline, with 10 μM nutlin or etoposide in the case of eGFP-p53 cells. Cells were washed in ice cold PBS, before being irradiated with 0.3 J/cm^2 of UV light. Cells were lysed immediately after irradiation by the addition of 1 ml ice cold iCLIP GFP lysis buffer to each plate. Lysates were collected using cell scrapers, and homogenised by passing through 5 ml syringes fitted with 27 G needles until sufficiently homogenised. Lysates were then centrifuged(17,900 g, 10 minutes, 4^{o}C), supernatants were collected, and input aliquots taken, before storage at -80^{o}C.

Preclearing of samples was performed by adding pre-equilibriated control agarose resin beads (25 μl bead slurry per sample) to the lysates, and incubating at 4^{o}C, with rotation, for 30 minutes. Samples were then centrifuged, and supernatants collected. GFP-trap agarose beads were added to each sample, and incubated at 4^{o}C, with

rotation, for 2 hours. Beads were then washed four times in room temperature high salt buffer, followed by two washes in ice cold PNK wash buffer. After the final wash, beads were collected and resuspended in 100 μl PNK buffer, allowing 10 μl to be taken for further analysis.

The remaining beads were collected by centrifugation, before being resuspended in 40 μl PK buffer + 10 μl Proteinase K (Roche). Samples were incubated at 37^{o}C, 1100 rpm for 20 minutes, before 50 μl PK-urea buffer was added, and the incubation step repeated. 100 μl phenol/chloroform (Invitrogen) was added to each sample, and incubated at 37^{o}C, 1100 rpm for 5 minutes. Samples were transferred to Maxtract high density eppendorfs (Qiagen), which had been centrifuged at full speed for 5 minutes at room temperature prior to use. Samples were incubated a second time at 37^{o}C, 1100 rpm for 5 minutes, before centrifugation (maximum speed, 5 minutes, room temperature). The upper aqueous phase was transferred to a low DNA binding tube (Eppendorf). RNA was then precipitated by adding 1.2 μl glycoblue (Invitrogen), along with 10 μl 3 M sodium acetate. The solution was mixed before adding 250 μl ice cold 100% ethanol, and precipitated overnight at -80^{o}C. To pellet RNA, samples were centrifuged (23,897 g, 20 minutes, 4^{o}C). The resulting supernatant was removed, and the RNA pellet washed with 80% ethanol, followed by a second centrifugation step (23,897 g, 5 minutes, 4^{o}C). Supernatant was discarded, and an additional centrifugation used (23,897 g, 30 seconds, 4^{o}C) to facilitate removal of residual ethanol. Pellets were air dried, before resuspension in 7 μl nuclease free water. RNA quality was examined by running 1 μl of RNA on a bioanalyser using an RNA pico chip (both Agilent), following the manufacturer's protocol.

Libraries were prepared using NEBNext Ultra II Directional RNA library Prep Kit for Illumina (NEB) according to manufacturer's protocol for purified mRNA or rRNA depleted RNA. In brief, RNA was fragmented for 15 min at 94^{o}C and then reverse transcribed. Second strand DNA synthesis was the performed, and the resulting double-stranded cDNA was purified with NEBNext sample purification beads (NEB). After end repair, adaptors were ligated and uracils excised using the USER enzyme. Samples were purified with the sample purification beads. PCR amplification is then used to enrich for adaptor ligated cDNA, whilst adding index primers. Libraries were amplified depending on the original RNA concentration of the library. A final library clean up was performed, before the quality of the libraries were assessed on a bioanalyser using a high sensitivity DNA chip (Agilent).

Libraries with an average length of 320 nt were pooled and sequenced with an Illumina NextSeq 10 instrument, using 81 nucleotide paired-end sequencing mode with a NextSeq 500/550 High 11 Output v2.5 kit (75 cycles, Illumina).

3.6 iCLIP

3.6.1 Generation of irCLIP linker

An infrared-dye-conjugated and biotinylated ligation adaptor was described by Zarnegar *et al.* (2016) [93]. The L3-IR linker (Table 3.8) was ordered from IDT with 5' phosphorylation. Pre-adenylation of the L3-IR linker 5' end was performed using a 5' DNA adenylation kit (NEB), according to the manufacturer's protocol, with the following reaction:

Reactions were incubated at 65^{o}C for 2 hours, followed by inactivation at 85^{o}C for

Table 3.8: L3-IR linker 5' end pre-adenylation.

Component	Volume
L3-IR linker (100 μM)	200 μl
10x 5' DNA Adenylation reaction buffer	75 μl
ATP (1mM)	75 μl
Mth RNA ligase	100 μl
Nuclease free water	300 μl

10 minutes. The reaction was then split in half, and each precipitated overnight at -20^{o}C by the addition of 40 μl 3M sodium acetate (pH 5.5) and 1 ml 100% ethanol. The L3-IR linker was then precipitated by centrifugation (21800 g, 30 minutes, 4^{o}C). The supernatant was removed, and pellets washed in 500 μl 80% ethanol, before resuspension in 180 μl PBS. 20 μl of 10 mM IRdye-800CW-DBCO (LI-COR) was then added, and reactions incubated at 37^{o}C for 2 hours. The two reactions were then pooled and mixed with 9.6 ml PNI buffer (Qiagen). 250 μl aliquots of the solution were then distributed across QIAquick nucleotide removal columns (Qiagen) and centrifuged (3824 g, 1 minute). Flow throughs were removed and the columns washed with 750 μl 80% ethanol, before centrifugation was repeated. Flow throughs were removed and an additional centrifugation performed to remove residual ethanol. Elution was peformed by adding 50 μl nuclease free water to each column, incubating at room temperature for 2 minute, and centrifuging (17,949 g, 1 minute). All eluates were then pooled and linker quantified using the nanodrop. The solution was diluted to 1 μM, aliquoted and stored at -20^{o}c.

3.6.2 iCLIP

In order to identify RNA binding sites at a high resolution, we employed iCLIP-seq (Huppertz *et al.*, 2014 [94]; König *et al.*, 2010 [95]) with a number of adaptations based on the eCLIP-seq procedure (Van Nostrand *et al.*, 2016 [96]). GFP and

eGFP-p53 Hek293 Flp-In T-REx cells were seeded in 3x 15 cm dishes, and once approximately 80% confluent, induced for 6 hours with the relevant stimuli. Cells were washed with PBS and UV irradiated with 300 mJ/cm^2 at 254 nm. Cells were then lysed with 1 ml of iCLIP GFP lysis buffer, and the three plates of each condition set were pooled. Lysates were then passed through a 27 G needle three times, and sonicated using a Digenonde bioruptor (level M, 4^{o}C, 10 seconds on, 15 seconds off, x3). The homogenate was centrifuged (17,900 g, 10 minutes, 4 ^{o}C), and supernatants taken, topped up to 2 ml with iCLIP GFP lysis buffer, and split into 1 ml aliquots. Each technical replicate used two of these 1 ml aliquots.

To degrade DNA, and obtain RNA fragments of the desired size, 1 ml of lysate was incubated with 20 U RNase I (Invitrogen) and 4 U Turbo DNase (Invitrogen) for 3 min at 37^{o}C, with 1100 rpm agitation. Lysates were then placed on ice and supplemented with 440 U RiboLock RNase Inhibitor (Thermo Fisher). 40 μl of control agarose bead slurry was pre-equilibrated in iCLIP GFP lysis buffer and incubated with the lysates for 30 min at 4^{o}C, with gentle rotation. The supernatants were then collected by centrifugation (2500 g, 4^{o}C, 2 minutes), before incubation with pre-equilibrated GFP-trap agarose beads (40 μl slurry) for 2 hours at 4^{o}C, with gentle rotation. Next, the beads were collected by centrifugation (2500 g, 4^{o}C, 2 minutes) and washed twice with 1 ml of high salt buffer, twice with 1 ml of medium salt buffer, and twice with 1 ml of PNK wash buffer. Beads were then resuspended in 20 μl PNK mix as follows.

Reactions were incubated for 20 minutes, at 37^{o}C with agitation (1100 rpm). Beads were then washed once with PNK wash buffer, once with high salt buffer, and twice with PNK wash buffer. Beads were resuspended in 20 μl ligation mix as

Table 3.9: iCLIP PNK mix.

Component	Volume
Nuclease free water	15 μl
5x PNK buffer	4 μl
Ribolock RNase inhibitor	0.5 μl
Polynucleotide kinase	0.5 μl

follows.

Table 3.10: iCLIP ligation mix.

Component	Volume
Nuclease free water	8 μl
4x Ligation buffer	5 μl
Pre-adenylated L3-IR linker (1 μM)	2 μl
PEG400	4 μl
Ribolock RNase inhibitor	0.5 μl
T4 RNA ligase	1 μl

Reactions were incubated overnight at 16^{o}C, with agitation (1100 rpm).

Beads were then washed with 500 μl of cold PNK wash buffer and three times with 1 ml of high salt buffer. Beads were transferred to a low binding tube during the third wash, and samples from the same replicate pooled. The beads were further washed twice with 1 ml ice-cold PNK wash buffer and resuspended in 20 μl PNK wash buffer, 1x NuPAGE loading buffer (Invitrogen) and 100 mM DTT, and denatured at 70^{o}C for 10 minutes with agitation (1200 rpm). The supernatant was collected by centrifugation (2500 g, 4^{o}C, 1 minute), and loaded on a 4-12% Bis-Tris NuPage gel (Invitrogen) and run for 90 minutes at 150 V in 1x MOPS running buffer (Invitrogen). Protein-RNA complexes were transferred to a nitrocellulose membrane (30 V for 1.5 hours). The region corresponding to the size of the protein of interest + 150-200 kDa was then cut out, transferred to a fresh microfuge tube, topped up with 200 μl of PK buffer and 800 μg of proteinase K, and incubated for 20 minutes at 37^{o}C with agitation (1100 rpm). Subsequently, 200 μl of PK-urea buffer was

added, and the sample incubated for 20 minutes at 37^{o}C with agitation (1100 rpm). Phenol/chloroform RNA extraction was then performed (Huppertz *et al.*, 6 2014; König *et al.*, 2010). Pellets were resuspended in 6 μl of nuclease free water, and stored at -20^{o}C. RNA quality was inspected by running 1 μl RNA on a bioanlyser RNA pico chip, following the manufacturer's protocol.

Reverse transcription was carried out using Superscript III (Invitrogen) and unique Rclip primers (Huppertz *et al.*, 2014; König *et al.*, 2010). Briefly, unique Rclip primers were added to each sample, and incubated at 70^{o}C for 5 minutes, before cooling to 25^{o}C. The reverse transcription reaction mix was then added as follows.

Table 3.11: iCLIP reverse transcription reaction mix.

Component	**Volume**
Nuclease free water	7 μl
5x First strand buffer	4 μl
DTT 0.1 M	1 μl
Ribolock RNase inhibitor	0.5 μl
Superscript III enzyme	0.5 μl

Reactions were then transferred to PCR tubes, and the following PCR cycling conditions used:

Table 3.12: iCLIP reverse transcription PCR cycling conditions.

Temperature	**Time (mm:ss)**
25^{o}C	05:00
42^{o}C	20:00
50^{o}C	40:00
80^{o}C	05:00
4^{o}C	∞

The reaction was then transferred to a low DNA binding tube and precipitated with ethanol (Huppertz *et al.*, 2014; König *et al.*, 2010). Pellets were resuspended in 12 μl of 1x TBE-urea loading buffer, heated for 3 min at 80^{o}C and separated

on a 6% TBE-urea precast gel (Invitrogen) for 40 min at 180 V. The region of gel containing cDNA fragments 80-200 nucleotides in length were cut from the gel. It is worth noting that high and medium sized fractions were combined together, as opposed to being processed seperately as in some protocols. 0.5 ml microtubes were pierced with a needle to form a hole in the bottom. Gel slices were transferred to this 0.5 ml microtube, which in turn was placed inside a 2 ml microfuge tube, and centrifuged (16000 g, 2 minutes), to break up the gel pieces thoroughly. 400 μl diffusion buffer was then added to the resulting gel, and incubated at 50^{o}C for 30 minutes, with agitation (1200 rpm). The disrupted gel was then filtered through a Costar SpinX column (Sigma), fitted with 2 glass prefilters (Whatman), by centrifugation at 16000 g. cDNA was then extracted using phenol/chloroform (Huppertz *et al.*, 2014; König *et al.*, 2010). Pellets were resuspended in the following ligation mix to enable circularisation of the cDNA, and transferred to PCR tubes.

Table 3.13: iCLIP ligation mix for circularisation.

Component	**Volume**
Nuclease free water	6.5 μl
10X CircLigase buffer II	0.8 μl
$MnCl_2$ 50 mM	0.4 μl
CircLigase II	0.3 μl

Reactions were incubated for 1 hour at 60^{o}C. Next, the cut oligo was annealed by adding the following reaction mix.

Table 3.14: iCLIP oligo annealing mix.

Component	**Volume**
Nuclease free water	25 μl
NEBuffer 4	4 μl
Cut oligo (10 μM)	1 μl

The following thermo programme was then used to enable annealing of the cut oligo.

Table 3.15: iCLIP oligo annealing conditions.

Temperature	Time (mm:ss)
95^{o}C	01:00
95^{o}C $\Rightarrow$ 25^{o}C	00:20 per 1^{o}C
25^{o}C	∞

The samples were then digested with 2 μl of FastDigest BamHI (Invitrogen) and incubated for 30 minutes at 37^{o}C, followed by 5 minutes at 80^{o}C. cDNA was then ethanol precipitated (Huppertz *et al.*, 2014; König *et al.*, 2010). Pellets were resuspended in 21 μl nuclease free water. Test PCRs were then performed as follows to optimise PCR cycling conditions for the final library amplification. Note PCR reaction mixtures are prepared in a different room to those in which the previous iCLIP steps were performed.

Table 3.16: iCLIP test PCR mix.

Component	Volume
Nuclease free water	3.75 μl
P5/P3 Solexa primer mix (10 μM each)	0.25 μl
Accuprime Supermix 1	5 μl
cDNA	1 μl

The libraries were then amplified using the following PCR cycling conditions, with a range of cycles to optimise library amplification.

Table 3.17: iCLIP test PCR cycling conditions.

Temperature	Time (mm:ss)	Cycles
98^{o}C	01:00	1
94^{o}C 65^{o}C 68^{o}C	00:15 00:30 00:30	20-30
68^{o}C	03:00	1
25^{o}C	∞	∞

PCR products were then analysed on a 6% TBE precast gel in TBE buffer. 10 μl PCR product was added to 2.5 μl 5x Hi-Density TBE loading buffer (Invitrogen) and loaded onto the gel along with a 1:10 dilution of low molecular weight DNA marker (NEB), and run for 60 minutes for 140 V. The gel was then stained with 1x TBE plus 1x SybrGold (Invitrogen) and visualised using the LICOR. If appropriate cDNA bands are visualised, final library amplification can be performed. The following reaction mix was PCR amplified using the validated number of PCR cycles. Note amplification of the remaining cDNA occurs in two batches.

Table 3.18: iCLIP final PCR mix.

Component	**Volume**
Nuclease free water	10 μl
P5/P3 Solexa primer mix (10 μM each)	1 μl
Accuprime Supermix 1	20 μl
cDNA	10 μl

PCR products were again mixed with 5x HiDensity TBE loading buffer and run on a 6% TBE precast gel in TBE buffer, along with a 1:10 low molecular weight DNA marker, for 180 V for 20 minutes. The gel was then stained using a mix of 1x TBE and 1x Sybrgold for 30 minutes, and visualised under a blue light transilluminator. Gel slices of the correct cDNA size were cut and added to 100 μl diffusion buffer, before incubation at 50^{o}C for 30 minutes, with agitation (1100 rpm). Dissolved gel was then filtered through glass prefilters and a Costar SpinX column as described previously. The resulting sample was then purified using the QIAquick gel extraction kit (Qiagen), and quantified on a bioanalyser using the DNA high sensitivity chip and reagents. Libraries were pooled for sequencing using single-end sequencing with a NextSeq 500/550 High Output V2 kit (75 cycles, Illumina).

3.7 Analysis

3.7.1 CLIP-seq analysis

Demultiplexed fastq files were aligned to the human genome using the STAR RNA-seq aligner [97]. Multimapping and spliced reads were removed using custom code. GFP-p53 sample signal was then compared to GFP control signal across the genome, and only regions with log2(ratio GFP-p53 over GFP) > 2 were retrieved, using bedtools [98]. Genomic features and biotypes were then counted using HTseq count [99]. Edge R [100] was then used to perform differential expression analysis on resulting read counts. Additionally, high confidence targets were identified by assessing presence of transcripts in all 3 replicates for each GFP-p53 treatment condition.

4

Is p53 an RNA binding protein?

4.1 Introduction

Previous *in vitro* works revealed that p53 can interact with RNA [87]. These studies utilised predominantly acellular approaches, utilising recombinant or heterologously expressed proteins in combination with *in vitro* transcribed RNAs. Whilst informative, *in vitro* interactions between basic proteins and RNA can occur in a non-specific, and non-physiological fashion, due to electrostratic interactions between the negatively charged phosphate backbone of nucleic acids, and highly positively charged regions on protein surfaces. Indeed, it is unsurprising that a protein with established nucleic acid binding capabilities, along with intrinsically disordered and basic regions, such as p53, could establish electrostatic interactions with the negatively charged RNA backbone under non-physiological conditions. Therefore, the relevance of such interactions must be tested within a cellular context. However, attempts to characterise the RNA binding activity of p53 in living cells have been fruitless to date. Yoshida *et al.* [74] used low stringency conditions to coimmunoprecipitate RNA and cytosolic p53 in MCF-7 cells, although this approach risks redistribution of macromolecules within complexes after cell lysis [101]. More importantly, the purification of protein complexes instead of individual proteins makes it impossible to confidently distinguish between direct and indirect interactions with RNA. Whilst *in vitro* approaches provide valid insights into the innate propensity of p53 to bind to RNA molecules, they offer little biological insight, or assessment of the physiological role and relevance, if any, of such interactions. This chapter aims to address the lack of *in vivo* evidence for p53-RNA interactions, whilst also generating a suitable system to characterise these interactions further.

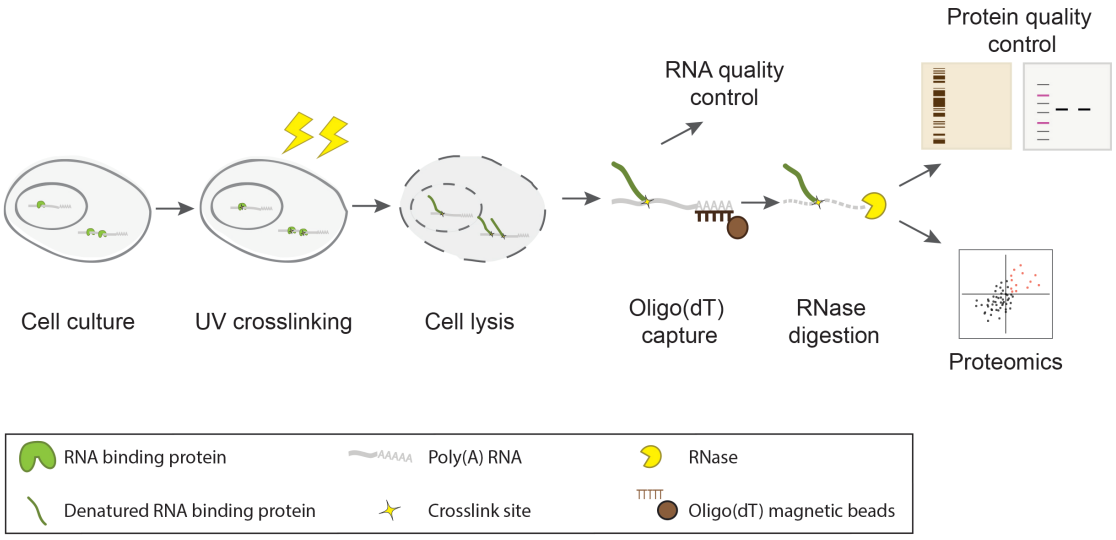


Figure 4.1: RNA interactome capture protocol. Schematic representation of the RNA interactome capture protocol. Cells are UV irradiated to induce covalent bonds between nucleotides and amino acids in direct contact. Cells are then lysed under denaturing conditions, and polyadenylated RNAs annealed to oligo(dT) beads. Stringent wash steps remove non-covalently bound interactors, before RNase treatment liberates RNA-bound proteins. These poly(A)-RNA binding proteins can then be identified by proteomic analysis.

4.2 p53 binds polyadenylated RNA in living cells

To determine whether p53 directly interacts with RNA in living cells, RNA interactome capture (RIC) was performed (Figure 4.1). Briefly, RIC utilises UV irradiation of cell monolayers to induce the formation of covalent bonds between naturally photoreactive nucleotide bases and amino acids, essentially freezing RNA and proteins in direct contact (termed zero distance). Oligo(dT) beads are used under denaturing conditions to capture polyadenylated (poly(A)) RNA-bound proteins. RNase digestion liberates bound proteins, which can be identified specifically via western blotting, or globally via mass spectrometry to produce cellular mRNA interactomes, also known as RBPomes [10, 9].

In this case, RIC was performed using MCF-7 cells, a human breast carcinoma derived cell line that expresses wild-type p53. As p53 is unstable under normal

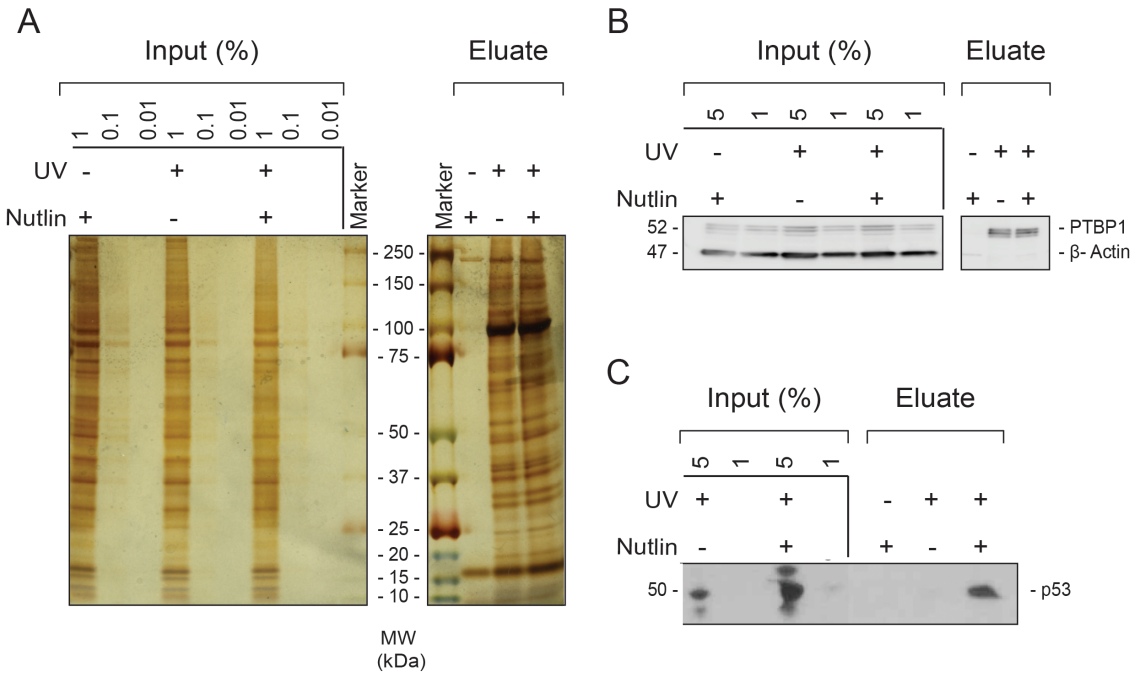


Figure 4.2: p53 binds polyadenylated RNA in living cells. RNA-IC performed on nutlin and DMSO treated MCF-7 cells. RNA-IC inputs and eluates were examined by silver staining (A) and western blotting (B) against RNA binding protein PTBP1, and non-RNA binding β-actin. Western blotting against p53 (DO-1) (C), demonstrates increased poly(A)-RNA bound p53 in nutlin treated, UV irradiated eluates.

conditions, we treated the cells with nutlin-3a, a small molecule antagonist of the p53-MDM2 interaction. Nutlin-3a binds specifically to the hydrophobic p53 binding pocket of the MDM2 protein a E3 ubiquitin ligase, resulting in disruption of the MDM2-p53 interaction [102]. As a result, subsequent autoregulatory feedback loops are inhibited, preventing ubiquitination and proteasomal degradation of p53 by MDM2, whilst also disrupting MDM2-mediated nuclear export of p53, and inhibitory binding of MDM2 to the p53 transactivation domain. As a consequence, nutlin treatment leads to stabilisation and accumulation of p53, subsequent activation of the p53 pathway and induction of the relevant downstream cellular responses (cell cycle arrest, senescence, and apoptosis) [102, 103].

MCF-7 cells were treated with nutlin (10 uM) or DMSO, for 6 hours. Non-UV-

irradiated nutlin treated samples were included (henceforth termed non-crosslinked, or NCL), as well as the UV-irradiated nutlin and DMSO treated samples. Inputs (whole cell lysate) were taken, before RIC was performed. The resulting inputs and eluates for each sample were analysed by gel electrophoresis combined with silver staining to demonstrate differences in the protein patterns of experimental and control samples (Figure 4.2 A). Importantly, a complex protein pattern was detected in RIC eluates only when UV was applied to cells. Indeed, the non-UV irradiated (NCL) control is devoid of proteins, except for low molecular weight bands corresponding to RNases. This confirms that RIC efficiently removes any protein that is non-covalently linked to RNA. Moreover, RIC eluates exhibit an altered protein pattern when compared to inputs, demonstrating that the RIC protocol captures a specific subset of proteins that correspond to previous RBPomes [10, 9]. Similarity between the proteins isolated from crosslinked RIC samples in presence or absence of nutlin suggests that dominant cellular RBPs are not affected by nutlin treatment.

Western blot analysis of inputs and eluates (Figure 4.2 B) shows enrichment for well characterised RBPs in eluates. For example, polypyrimidine tract binding protein 1 (PTBP1) - a splicing factor that binds intronic polypyrimidine repeats – is only detected if UV is applied to cells. In contrast, the non-RNA binding beta-actin is only observed in input material, and is thus fully depleted in all RIC eluates. These two examples shows that eluates are enriched for RBPs, and depleted of non-RNA binding proteins, thus verifying the high stringency of the approach in MCF-7.

p53 is detected at varying levels in both nutlin treated and DMSO control in-

puts, with nutlin treatment causing an evident accumulation of p53 (Figure 4.2 C). Importantly, p53 is only detectable in nutlin treated, cross-linked eluates, and is absent in DMSO treated and NCL eluates. This demonstrates that upon stabilisation by nutlin, p53 interacts with RNA in living cells and is isolated by RIC, even in the presence of high salt and chaotropic detergents. Taken together, these results confirm that p53 is a bona fide RBP in cellulo.

4.3 A lack of specificity of p53 antibodies presents challenges

After confirming that p53-RNA interactions occur in living cells, my goal was to establish a suitable system with which to further explore these interactions, with the ultimate aim to identify p53 RNA-binding specificity and preferences in terms of transcripts and sequence. Previous attempts to identify and characterise p53-RNA interactions in living cells utilised anti-p53 immunoglobulin G (IgG) monoclonal antibodies to immunoprecipitate (IP) p53, before analysis of coimmunoprecipitated RNA species, either without crosslinking [104, 74], or after formaldehyde or UV crosslinking [104]. These approaches previously yielded little success due to the presence of contaminating ribonucleoprotein complexes when conventional IgG antibodies were used, even in p53 null cell lines [104, 87]. For example, I tested a broadly used p53 antibody (DO-1), targeting amino acid residues 11-25 in the N-terminus of the protein. DMSO and nutlin treated MCF-7 cells were subject to IP using the DO-1 antibody coupled to magnetic beads, using near-physiological (150mM) salt concentrations. IP inputs and eluates were subject to SDS-PAGE and silver staining

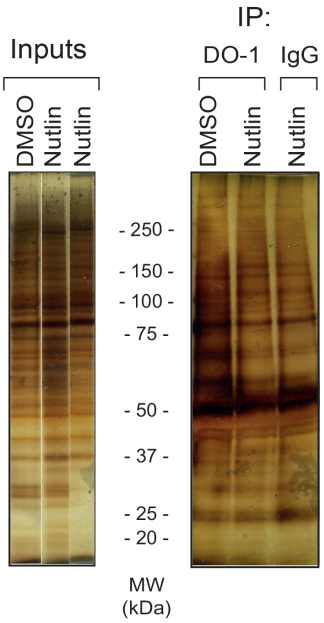


Figure 4.3: Immunoprecipitation using DO-1 antibody. Along with an IgG only control, DO-1 antibody was used to immunoprecipitate p53 in DMSO and nutlin treated MCF-7 cells. The resulting inputs and IP eluates were analysed by SDS-PAGE and silver staining to assess the presence of potential contaminating coimmunoprecipitating proteins.

(Figure 4.3). Eluate protein patterns exhibit multiple coimmunoprecipitating protein species, even in the IgG only control sample. These results recapitulated those previously described by Riley *et al.* [104], in which IgG antibodies may coimmunoprecipitate non-specific or contaminating RNP species. The protein complexity of these IPs means it is extremely difficult to perform any further successful protein purification based approaches, as discrimination between true interactions and contaminants becomes extremely challenging. In an effort to overcome this issue, single domain antibodies or nanobodies targeting p53 were used.

Nanobodies are derived from a specific subset of Camelid species' IgG, which are termed heavy chain only IgG, and lack light chains. These heavy-chain-only antibodies contain two constant domains (C_H2, C_H3), and only one variable, antigen binding domain (V_HH). Recombinant V_HH domains, or nanobodies, that are cloned

and expressed separately from the constant regions, have a mass of approximately 15 kDa, and exhibit highly specific antigen binding, stability and affinity [105]. Anti-p53 nanobodies pre-coupled to agarose beads (henceforth termed p53-trap) are commercially available (Chromotek), and are available in two forms, targeting either the N-terminus or C-terminus of p53 (termed N-t and C-t, respectively). General consensus from previous works is that the C-terminal domain harbours p53's RNA binding activity, and as such antibodies targeting N-terminal epitopes were preferred. Both N-t and C-t p53-trap were used here, in order to compare whether targeting the different epitopes resulted in differences in p53 IP efficiency and yields.

IPs were performed using nutlin treated MCF-7 cells, along with DMSO treated controls, utilising either N-t or C-t p53-trap, and high (500 mM) or physiological (150 mM) NaCl in the lysis and wash buffers. Increasing NaCl concentration to 500 mM will help reduce ionic and electrostatic attractions, with increased capacity to disrupt complexes, which may in turn reduce the level of non-specific contaminants. Inputs (whole cell lysate) and IP eluates were analysed by gel electrophoresis combined with silver staining and western blot (Figure 4.4). Silver staining shows a difference in protein band pattern between inputs and IPs (Figure 4.4 A). However, the IP eluates contain multiple protein bands, which may correspond to p53 protein partners, or contaminants. The incidence of co-precipitated species does decrease as salt concentration increases, suggesting that they are not p53 interactors but proteins recognised by the nanobodies or non-specifically bound to the agarose beads. In spite of this result, western blot analysis shows effective IP of p53 in eluates compared to inputs (Figure 4.4 B).

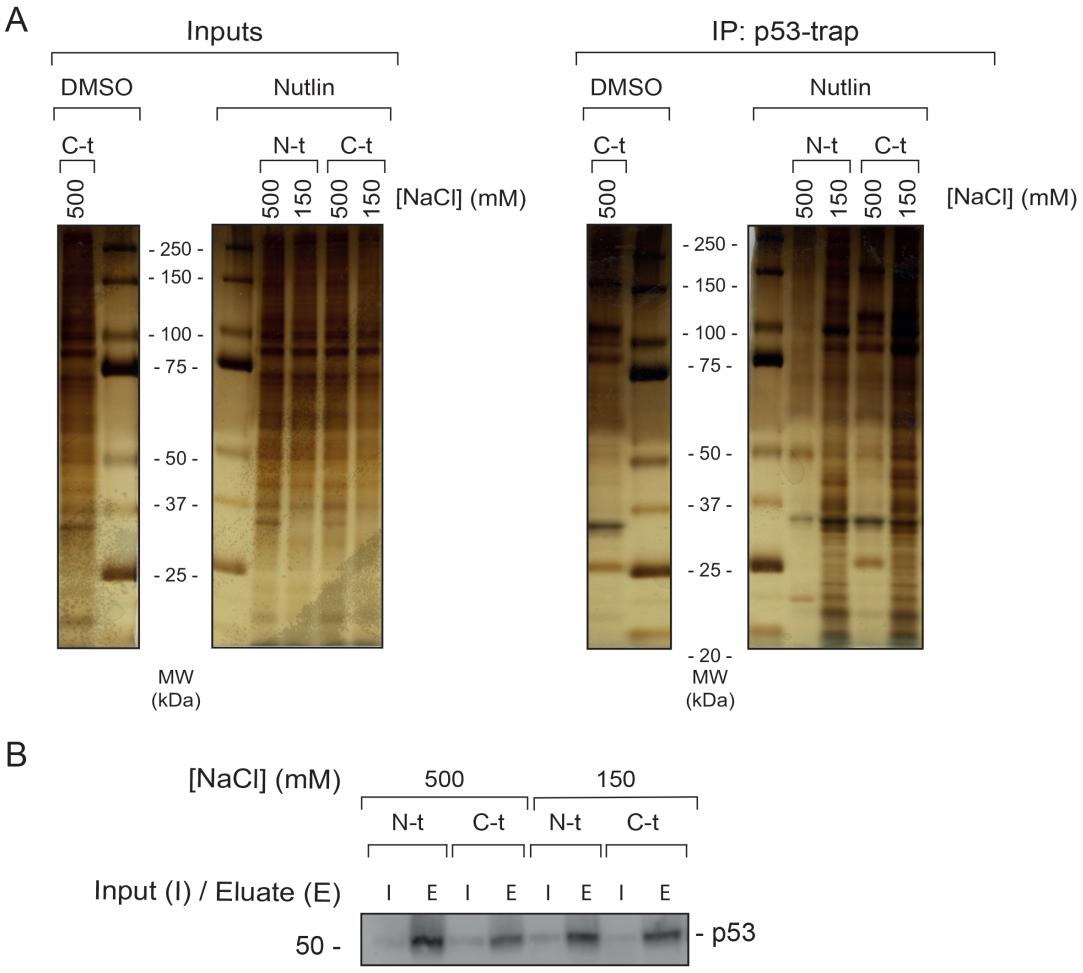


Figure 4.4: Immunoprecipitation using p53-trap nanobodies. p53-trap nanobodies targetting either N (N-t) or C-terminus (C-t) of p53 were used to immunoprecipitate p53 from DMSO or nutlin treated MCF-7 cells. High (500 mM) and low (150 mM) NaCl concentrations were used to assess whether IP stringency could be improved. Inputs and IP eluates were analysed by SDS-PAGE and silver staining (A) or western blotting (B) against p53 (DO-1).

To further investigate the identity and composition of p53-trap IP eluates, I performed an IP (150 mM NaCl) with either C-t (Figure 4.5 A) or N-t (Figure 4.5 B) p53-trap in samples treated with DMSO or nutlin. IP eluates were then analysed by label-free mass spectrometry (MS) (LC-MS and initial analysis performed by Svenja Hester). Our proteomic analysis revealed a high number of co-immunoprecipitating proteins, as demonstrated by the 'cloud' of proteins around the centre of each MA plot. Importantly, hundreds of co-precipitating proteins showed a higher nutlin/DMSO fold change and intensity than p53, which was expected to be a dominant protein in the IP. Of the 254 (C-t) and 389 (N-t) proteins identified as enriched in nutlin/DMSO, only 33 are shared between the two datasets. This low level of overlap further emphasises the lack of specificity observed in this system. Together, these results suggest that the N-t and C-t p53-trap system retains some of the problems highlighted when using conventional IgG antibodies, and suffer similar limitations in terms of differentiating between direct p53 interactors, and those representing non-specific/contaminant species.

4.4 An inducible GFP-p53 system increases specificity

To circumvent the limitations imposed by available p53 antibodies, I generated FLAG-HA and eGFP tagged p53 constructs, and transfected them into Hek293 Flp-In T-REx cells (wild type p53) to generate stable doxycycline-inducible cells lines. Due to the current consensus in the literature that p53 binds RNA through a portion of its C-terminal domain, the remaining work was performed using an

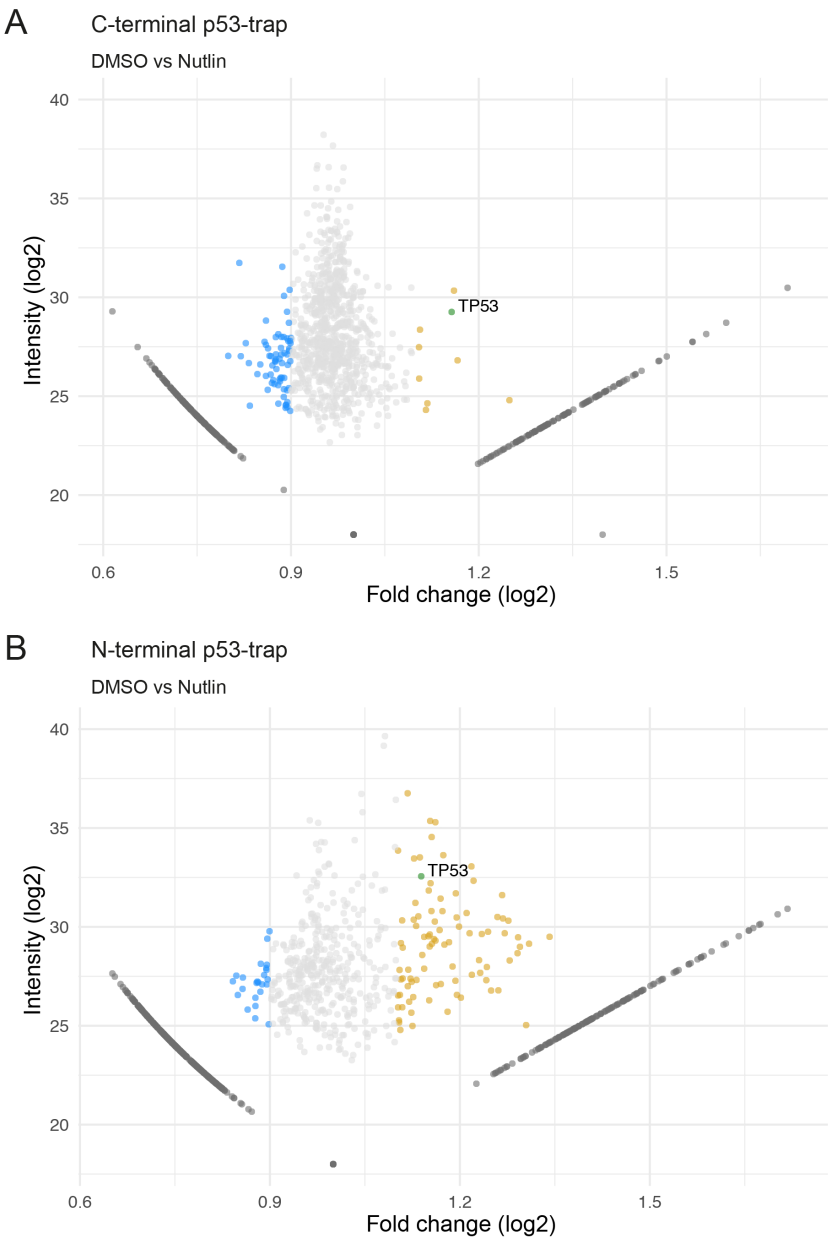


Figure 4.5: Proteomic analysis of p53-trap IPs. MCF-7 cells were treated with either DMSO or nutlin. p53 was then immunoprecipitated using p53-trap N-terminal and C-terminal targetted nanobodies under low (150 mM) NaCl conditions. Yellow and blue points signify proteins with a log2(FC) >1 or < 1, respectively. Dark grey points signify proteins that had an intensity of 0 in either condition. (A) C-terminal p53-trap IP. Log2(fold change) nutlin over DMSO treated cells. (B) N-terminal p53-trap IP. Log2(fold change) nutlin over DMSO treated cells.

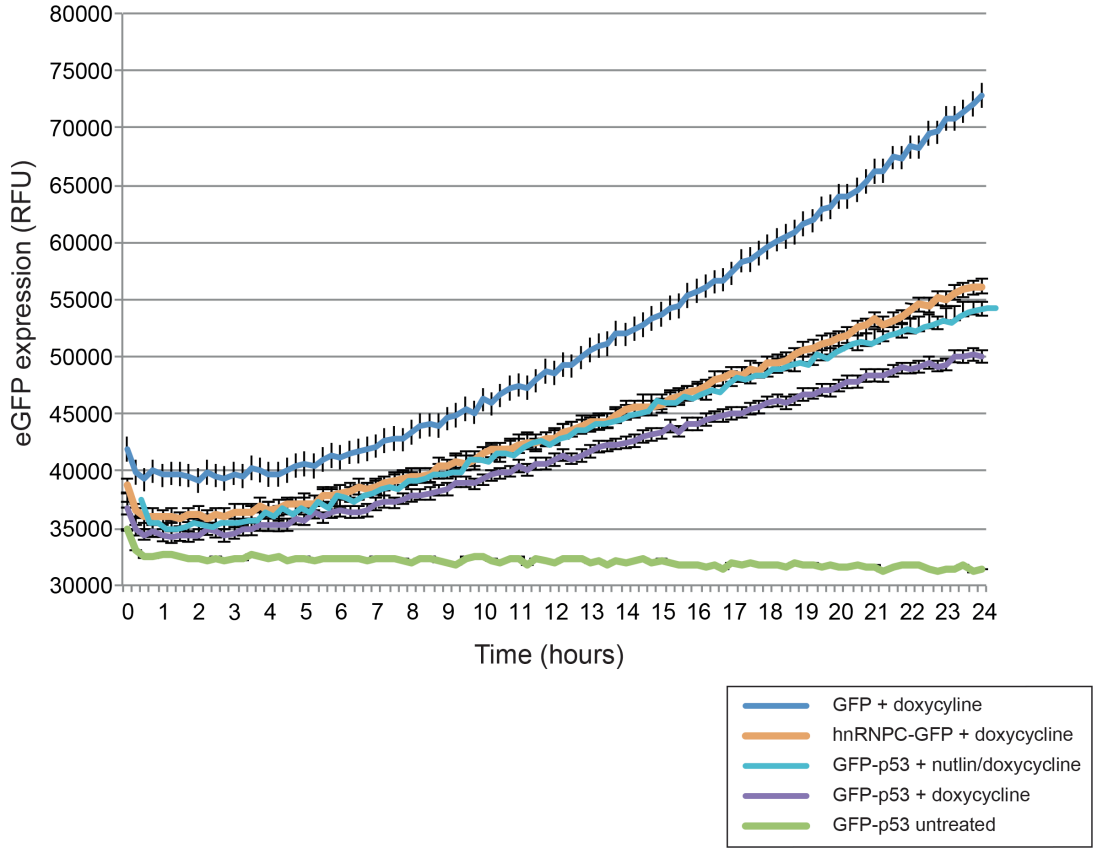


Figure 4.6: Assessing GFP expression in GFP-p53 Flp-In T-REx cells. Hek293 Flp-In T-REx expressing either no construct (termed parental), GFP, hnRNPC-GFP or GFP-p53 were treated with the indicated inducing conditions. Cells were incubated over a 24 hour period in a platereader with atmospheric control (37^{o}C with 5% CO_2) and eGFP measurements taken every 15 minutes. RFU denotes relative fluorescence units. Error bars show standard deviation across 3 replicates.

N-terminally tagged eGFP-p53 cell line (henceforth termed GFP-p53).

To ascertain whether the GFP-p53 cell line was inducible, and establish suitable induction conditions, eGFP fluorescence was monitored over a 24 hour period, using a temperature and atmospheric controlled plate-reader, with readings taken every 15 minutes. Figure 4.6 shows the relative fluorescence of GFP-p53 Hek293 cells treated with a variety of stimuli, along with GFP and hnRNPC-GFP control cells. Treatment of cells with doxycycline (henceforth termed dox), resulted in significant inductions of the control GFP and hnRNPC-GFP, along with GFP-p53. Treatment

with DMSO and nutlin has no impact on GFP-p53 fluorescence levels over time. Interestingly, addition of nutlin in combination with doxycycline appears to increase the level of GFP-p53 induction, with similar levels to that of the hnRNPC-GFP positive control. This could be explained by nutlin-mediated inhibition of Mdm2's negative regulatory effects on p53, resulting in enhanced stabilisation of the GFP-p53.

To validate the results of the plate reader assay, induction of eGFP-p53 and endogenous p53 was then analysed using western blotting. DMSO treated cells were used as a control, with little to no induction of GFP labelled or endogenous p53 (Figure 4.7 A), and potential depletion of both species over the 6 hour timeframe. Treatment with doxycycline (Figure 4.7 B) caused a substantial induction of GFP-p53, but had little to no impact on endogenous p53 levels. Treatment with nutlin (Figure 4.7 C) has limited impact on the GFP-p53 levels potentially increasing levels slightly up to 4 hours. Endogenous p53, on the other hand, is stabilised and accumulates over the 6 hour period, as anticipated. Combination of nutlin/doxycycline (henceforth nutlin/dox), results in significant induction of the GFP-p53 protein (Figure 4.7 D), confirming the results obtained with the platereader (Figure 4.6).

Once suitable conditions for induction of the GFP-p53 were established, the next step was to optimise IP conditions. An anti-GFP nanobody coupled to agarose beads (termed GFP-trap, Chromotek), proven to be extremely clean and efficient for both protein-protein and protein-RNA interactions [1, 106], was used. GFP-p53 was either uninduced, or induced using nutlin/doxycycline for 6 hours. A GFP only control sample was also used, which was induced for 6 hours using doxycycline. IPs were performed using physiological (150 mM) NaCl in the lysis and wash buffers. IPs

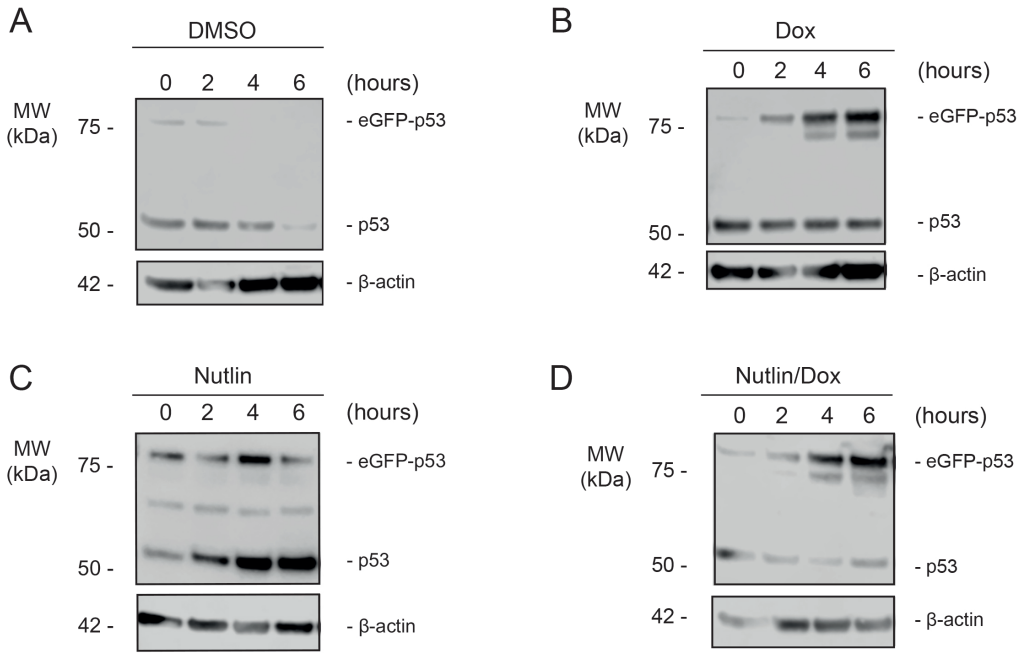


Figure 4.7: Assessing GFP-p53 expression in GFP-p53 Flp-In T-REx cells. Immunoblots of Hek293 GFP-p53 cells treated with DMSO (A), doxycycline (B), nutlin (C) or nutlin and doxycycline (D) over a 6 hour timecourse, blotting with anti-p53 DO-1 antibody. Doxycycline was used at 1 μg/ml, and nutlin at 10 μM).

were RNase digested after the first IP wash step, to remove RNA-mediated protein interactions. The resulting inputs and IP eluates were analysed by silver staining (Figure 4.8). Silver staining shows a marked difference in protein band pattern between inputs and eluates. The major bands in the GFP-trap IP correspond in size to GFP-p53 (approximately 75 kDa) and endogenous p53 (approximately 53 kDa), suggesting GFP-p53 is able to oligomerise with the endogenous protein. Even under physiological salt conditions, this IP approach shows substantial improvement over IgG antibodies against endogenous p53, in terms of signal-to-noise ratio, with excellent removal of contaminants.

p53 is able to co-ordinate a myriad of specific responses to a range of cellular stressors. As such, it was important to consider alternative p53 activating conditions, in addition to p53 stabilisation by nutlin. Accordingly, I then tested the GFP-p53

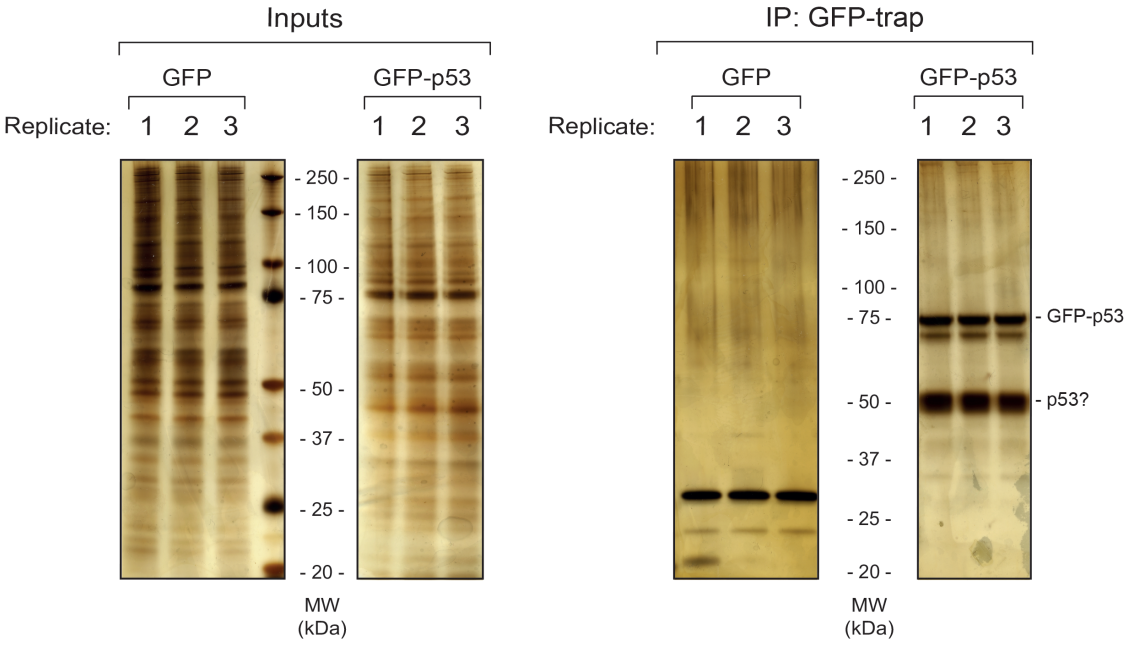


Figure 4.8: Immunoprecipitation using GFP-trap nanobodies. GFP-trap nanobodies were used to immunoprecipitate GFP and GFP-p53 in doxycycline and nutlin/doxycycline treated cells, respectively, using low (150 mM) NaCl in lysis and wash buffers. Inputs and IP eluates were analysed by silver staining to assess the presence of potential contaminating coimmunoprecipitating proteins.

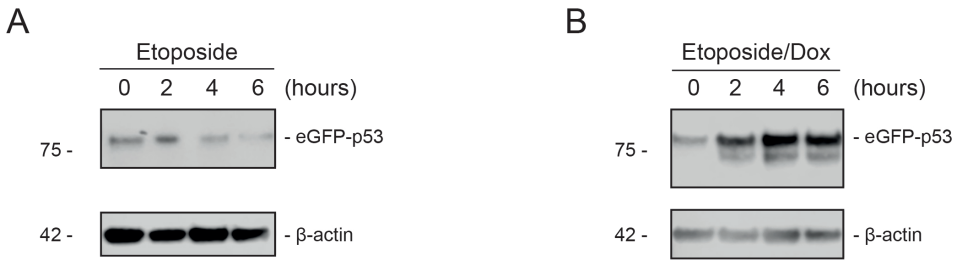


Figure 4.9: Assessing GFP-p53 induction upon DNA damage. Immunoblots of Hek293 GFP-p53 cells treated with etoposide (A), and etoposide plus doxycycline (B) over a 6 hour timecourse. Doxycycline was used at 1 μg/ml, and etoposide at 10 μM).

inducible system using etoposide treatment. Etoposide is a chemotherapeutic agent used to treat a range of cancers, including germ cell tumours, Hodgkin and Burkitt lymphomas , Ewing sarcomas and non-small cell lung cancers [107]. Etoposide targets type II topoisomerases (TopoII) which regulate DNA topology during cellular processes such as DNA replication, repair, transcription and the remodelling of chromatin, via the introduction of transient double strand breaks. Etoposide essentially 'poisons' the TopoII enzyme by stabilising the TopoII in complex with transient double-stranded DNA breaks, forming permanent protein-associated DNA breaks. As these DNA breaks accumulate, cellular machinery recognises this DNA damage, and acts to induce apoptosis to prevent proliferation of these potentially neoplastic cells [108, 107, 109]. Etoposide has previously been shown to activate p53 protein, likely via post-translational modifications [108], and induce downstream apoptotic pathways utilising both transcriptional-dependent and independent pathways [110, 111, 112]. GFP-p53 cells were treated with either etoposide alone (Figure 4.9 A) or etoposide combined with dox (Figure 4.9 B). Etoposide alone has little impact on GFP-p53 levels, whilst etoposide/dox treatment results in substantial induction of GFP-p53, from as early as 2 hours.

I then assessed whether GFP-p53 IPs performed after etoposide/dox treatment matched those obtained with nutlin/dox treated samples. GFP-trap IPs were performed on GFP and GFP-p53 samples 6 hours post induction with dox and etoposide/dox, respectively. Near physiological salt conditions (150 mM) were used. As previously, RNase digestions were performed after the first IP wash step. Inputs and IP eluates were examined using silver staining (Figure 4.10 A) and western blotting with an anti-GFP antibody (Figure 4.10 B). As with nutlin/dox treated samples, silver staining shows a marked difference in protein band pattern between inputs and eluates, with species at approximately 75 kDa and 50 kDa. Western blotting reveals enrichment for GFP-p53 in the IP compared to inputs. This validates the suitability of this system, using multiple induction conditions. Compared to previous IgG approaches, this enables significant enrichment of the target protein, whilst affording improved signal to noise ratios, and removing previously high levels of background/contaminating protein species.

To validate the enhanced biochemical tractability of this system, GFP-trap IP eluates obtained for each GFP-p53 induction condition, and for the GFP only controls, described above, were subjected to label-free proteomic analysis. Protein intensity ratios were calculated between GFP control and GFP-p53 samples, and the significance of these ratios estimated using moderated t-test. (LC-MS and initial analysis performed by Wael Kamel).

MA plots showing log2 fold change (FC) against average intensity (log2) in GFP-p53 samples, for nutlin/dox (Figure 4.11 A) and etoposide/dox were plotted (Figure 4.11 B). When compared to MA plots obtained using the endogenous p53-trap IP system (Figures 4.5), there is a substantial increase in the number of proteins en-

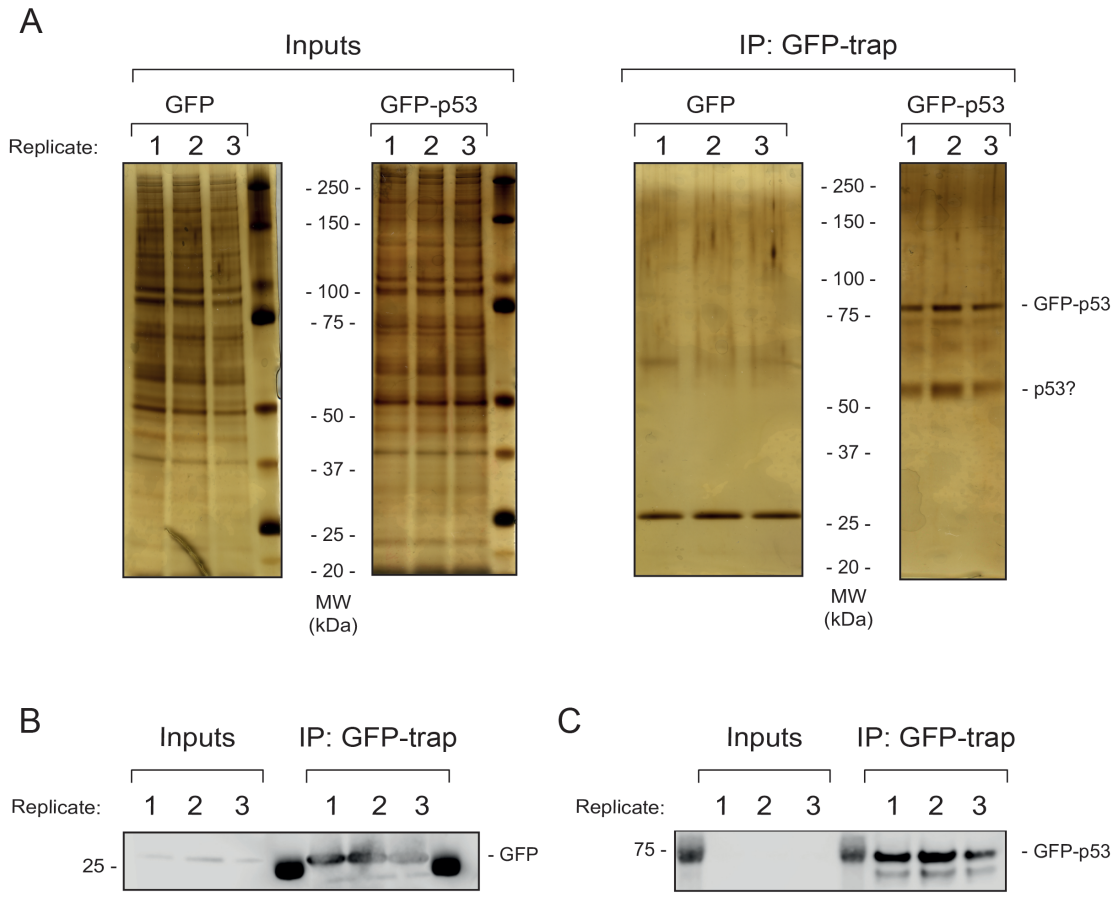


Figure 4.10: Immunoprecipitation using GFP-trap nanobodies. GFP-trap nanobodies were used to immunoprecipitate GFP and GFP-p53 in doxycycline and etoposide/doxycycline treated cells, respectively, using low (150 mM) NaCl in lysis and wash buffers. Inputs and IP eluates were analysed by silver staining (A) and western blotting (B) with an anti-GFP antibody (3H9).

riched in the target IPs over controls (right hand side of the plots), indicating that the IP efficiently enriches for specific interactors. Additionally, the 'cloud' of proteins around the centre of the plot is far smaller, suggesting lower incidence of background proteins. p53 is highlighted in green in each plot, and exhibits the highest intensity in GFP-p53 samples treated with either combination of stimuli, demonstrating the efficiency and stringency of this p53 pulldown approach.

Volcano plots show similar trends, plotting log2 fold change (FC) in GFP-p53 versus GFP cells against significance (-log10 adjusted P value) for nutlin/dox (Figure 4.12 A) and etoposide/dox (Figure 4.12 B) treated samples. Significant proteins are again highlighted. GFP-p53 IPs, with either induction condition, demonstrate significantly higher number of proteins than the GFP only control, with 402 significantly enriched proteins in the nutlin/dox IPs and 253 significantly enriched proteins in the etoposide/dox IPs. p53 is highlighted in green, and is one of the most significantly enriched proteins, with a logFC of 12.96 and an adjusted P-value <0.001 in both conditions.

4.5 eGFP-p53 interacts with known p53 interactors and RBPs

Intersection of the two GFP-p53 immunoprecipitation datasets shows a large number of common proteins coimmunoprecipitated under both induction conditions (Figure 4.13). 241 common proteins were identified (see Supplementary Table 1), demonstrating a high level of concordance between the two datasets.

Whilst this system is obviously useful in terms of its biochemical tractability,

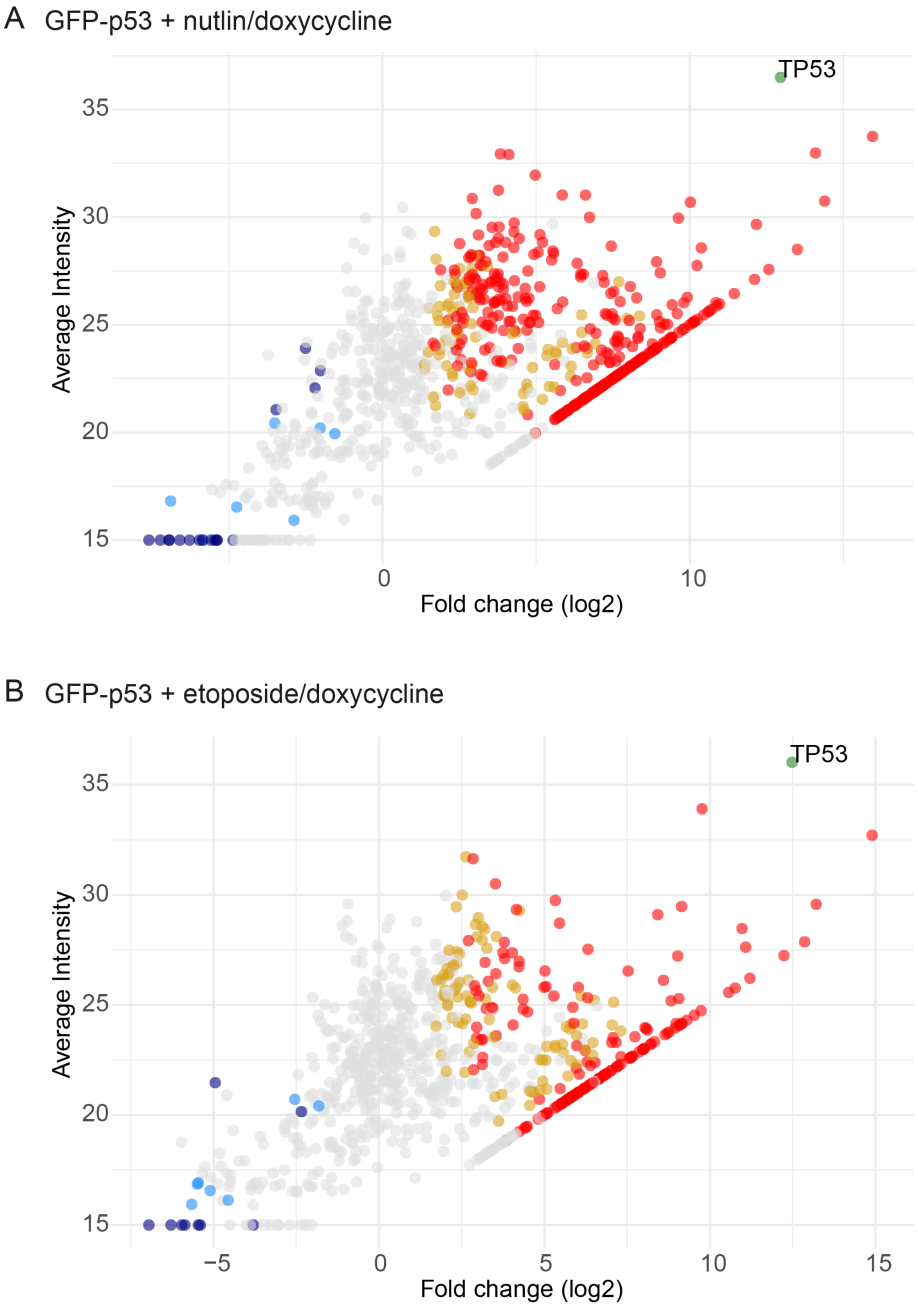


Figure 4.11: Proteomic analysis of GFP-p53 IPs. GFP-trap IPs were performed on induced GFP and GFP-p53 cells, using low (150 mM) NaCl conditions in order to investigate GFP-p53 protein interactions. Mass spectrometry and subsequent initial analysis was performed by Wael Kamel. MA plots show fold change in GFP-p53 versus GFP, plotted against average GFP-p53 intensity, for nutlin/doxycycline (A) and etoposide/doxycycline (B) treated cells. Yellow and light blue denote proteins with an adjusted P-value<0.05, whilst red and dark blue points denote proteins with an adjusted P-value<0.01.

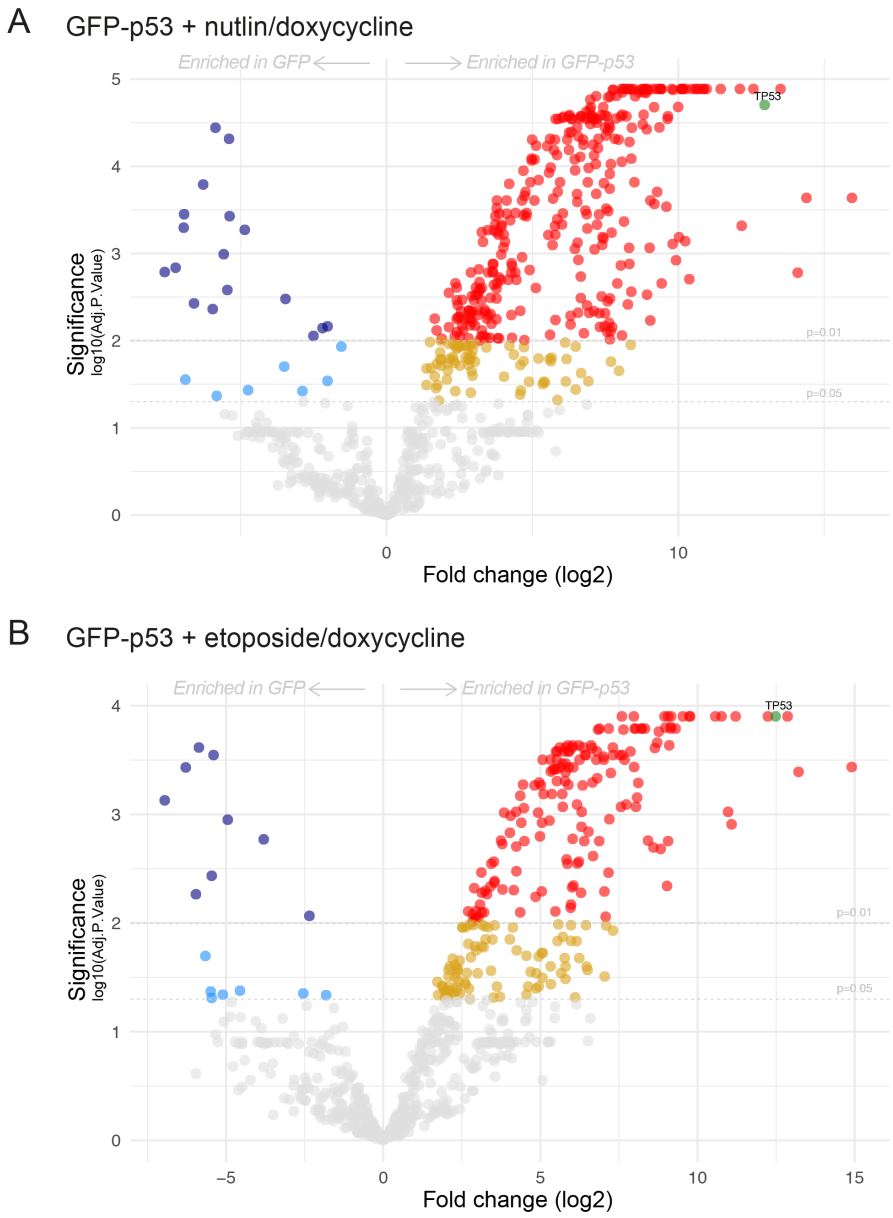


Figure 4.12: Proteomic analysis of GFP-p53 IPs. GFP-trap IPs were performed on induced GFP and GFP-p53 cells, using low (150 mM) NaCl conditions in order to investigate GFP-p53 protein interactions. Mass spectrometry and subsequent initial analysis was performed by Wael Kamel. Volcano plots show fold change in GFP-p53 versus GFP, plotted against adjusted P-value, for nutlin/doxycycline (A) and etoposide/doxycycline (B) treated cells. Yellow and light blue denote proteins with an adjusted P-value<0.05, whilst red and dark blue points denote proteins with an adjusted P-value<0.01.

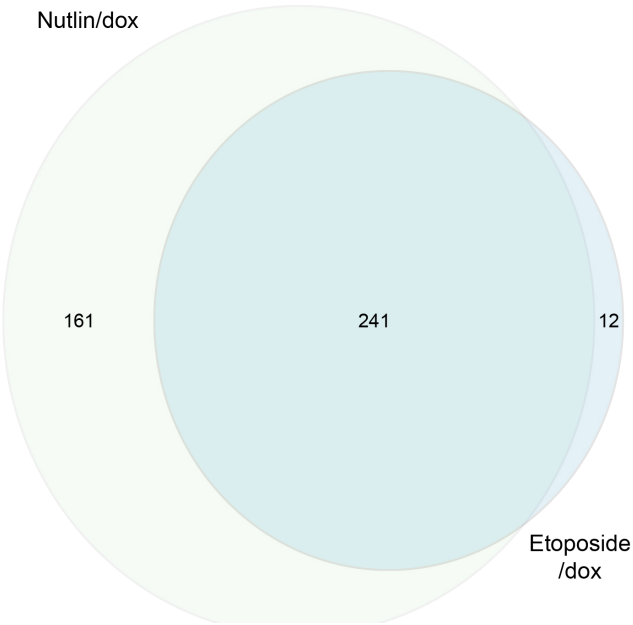


Figure 4.13: Commonly co-immunoprecipitated proteins. GFP-trap IPs were performed on induced GFP and GFP-p53 cells, using low (150 mM) NaCl conditions in order to investigate GFP-p53 protein interactions. Datasets for nutlin/doxycycline and etoposide/doxycycline treated cells were then intersected to find commonly immunoprecipitated proteins. The full list of shared proteins can be seen in Supplementary Table 1.

it was important that the system retained physiological relevance. To ascertain whether this is the case, I evaluated whether known p53 interactors were coimmunoprecipitated using this system, which would increase confidence in its biological applicability. The STRING protein-protein association database [113] was used to identify direct p53 interaction partners in each dataset (based on experimentally validated interactions, or those present in databases). Interaction partners over the significance threshold (adjusted P value < 0.05) are highlighted in volcano plots for each condition (Figure 4.14 A for nutlin/dox, and Figure 4.15 A for etoposide/dox), with each point's colour representing its STRING interaction score. STRING analysis of the resulting significant p53 protein interactome was then performed for each condition (Figures 4.14 B and 4.15 B). p53 itself is highlighted in green, with any known direct p53 interactors highlighted in blue, and previously unknown or indirect

p53 interactors coloured grey. Relevant complexes, or functional groups of proteins, are annotated.

54 previously characterised p53-protein interaction partners were identified in the set of common proteins, with 83 known interactors identified in the GFP-p53 nutlin/dox dataset, and 63 in the etoposide/dox dataset. This includes several proteins you would anticipate, including the key tumour suppressor and p53 regulator CDKN2A, and multiple components of the ubiquitin-proteasome machinery, involved in the regulation of p53 under normal cellular conditions, such as the E3 ubiquitin ligases CUL7 and CUL9, their activator NEDD8 and multiple components of the 26S proteasome. Interestingly, MDM2, which itself is an E3 ubiquitin ligase, is identified as a significant interaction partner in etoposide/dox treated cells, but falls below the significance threshold in the nutlin/dox samples (adjusted P value $<$ 0.1), likely due to the inhibition of this interaction by nutlin.

Several p53 interactors detected in Nutlin/dox IPs did not pass the statistical threshold in etoposide/dox IPs, including the p53 family member TP73, and TP53BP1. TRIM28, TRIM27, HDAC1 and HDAC2 are also significant hits in the nutlin/dox IPs. These proteins are all involved in the regulation of p53 proteins levels, and targeting p53 for proteasomal degradation. The presence of multiple p53 regulatory complexes and pathways across the two datasets supports the biological relevance and suitability of this system.

Protein complexes co-immunoprecipitated with p53 are annotated in figures 4.14 B and 4.15 B. Several RNA-associated complexes are highlighted, including multiple hnRNP proteins, splicing factors and ribosomal proteins. Of note is RPL5, which has previously been described to partake in the formation of an RPL5-MDM2-p53

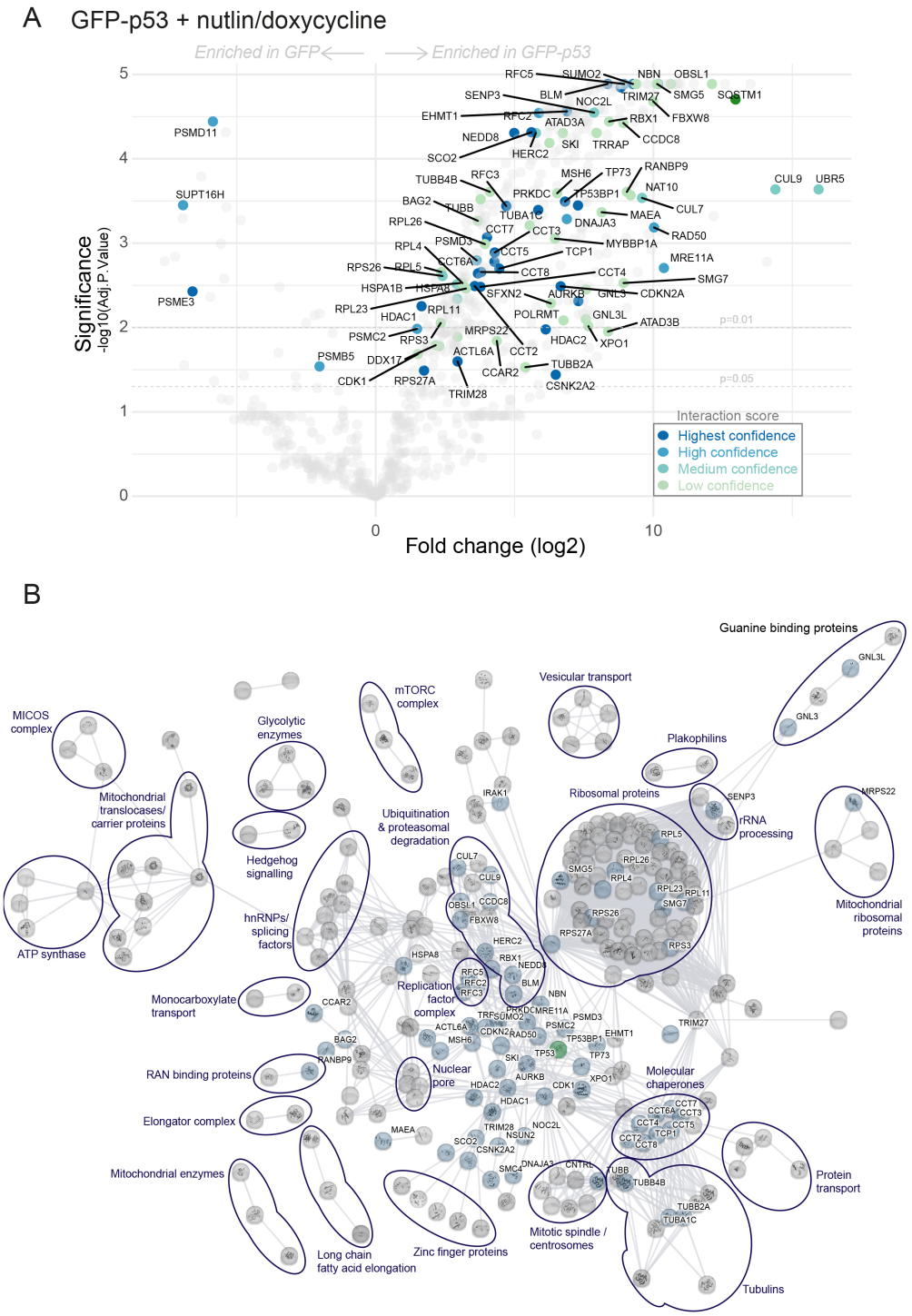


Figure 4.14: Nutlin/doxycycline - p53 protein interaction network. Volcano plots (A) show fold change in GFP-p53 versus GFP IPs, plotted against adjusted P-value, as previously, but with known p53 interactors highlighted based on confidence level (from STRING). STRING analysis of proteins enriched in GFP-p53 IP (Adj.P-value < 0.05) (B) shows known interactors highlighted in blue. Functional complexes are annotated.

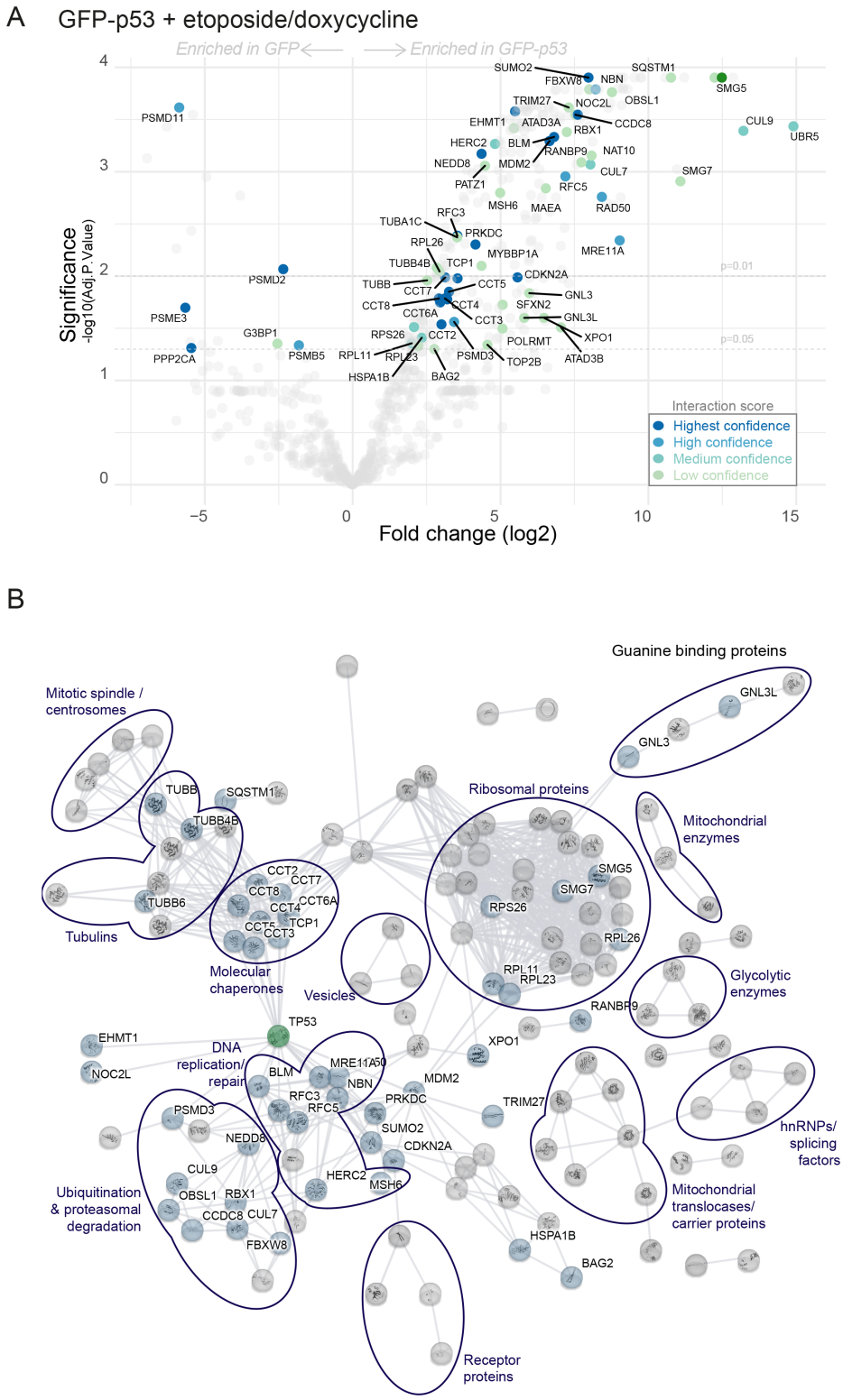


Figure 4.15: Etoposide/doxycycline - p53 protein interaction network. Volcano plots (A) show fold change in GFP-p53 versus GFP IPs, plotted against adjusted P-value, as previously, but with known p53 interactors highlighted based on confidence level (from STRING). STRING analysis of proteins enriched in GFP-p53 IP (Adj.P-value < 0.05) (B) shows known interactors highlighted in blue. Functional complexes are annotated.

complex, which was shown to bind the 5S and 5.8S rRNA [114], forming a potential p53-containing RNP complex [115]. Indeed, work by Brannan *et al.* [116], suggests that the likelihood of a protein being an RBP is related to the number of RBPs it interacts with. Figure 4.16 shows the number of RBPs present in each of the GFP-p53 IP conditions, with RBPs in red and blue denoting those with an adjusted P value < 0.05 and > 0.05, respectively. 254 RBPs are identified in the complete nutlin/dox GFP-p53 dataset, and 251 RBPs in the etoposide/dox GFP-p53 dataset, 70 of which are shared between the two. The presence of multiple hnRNPs, splicing factors, ribosomal proteins and eukaryotic initiation factors suggest p53 may associate with mRNAs undergoing translation. Indeed, the presence of polyadenylate-binding proteins 1 and 4 in this dataset act to support this idea, and could explain why p53 is isolated with polyadenylated RNAs, as evident in my RIC result (Figure 4.2). Overall, the identification of hundreds of RBPs associated with p53 supports p53 itself being an RBP, with potential links to mRNA metabolic processes.

Together, the data obtained from GFP-trap based purification of GFP-p53 and associated complexes, confirms that the GFP-p53 Hek293 Flp-In T-REx system enables improved biochemical tractability, whilst still offering reproducibility and a physiologically relevant context with which to study p53-RNA interactions. After establishing the system, and optimising induction of the GFP-p53 protein, optimisation of IP of GFP-p53 and interacting protein complexes was achieved. Next, I aimed to establish whether I was able to verify p53-RNA interactions within this system.

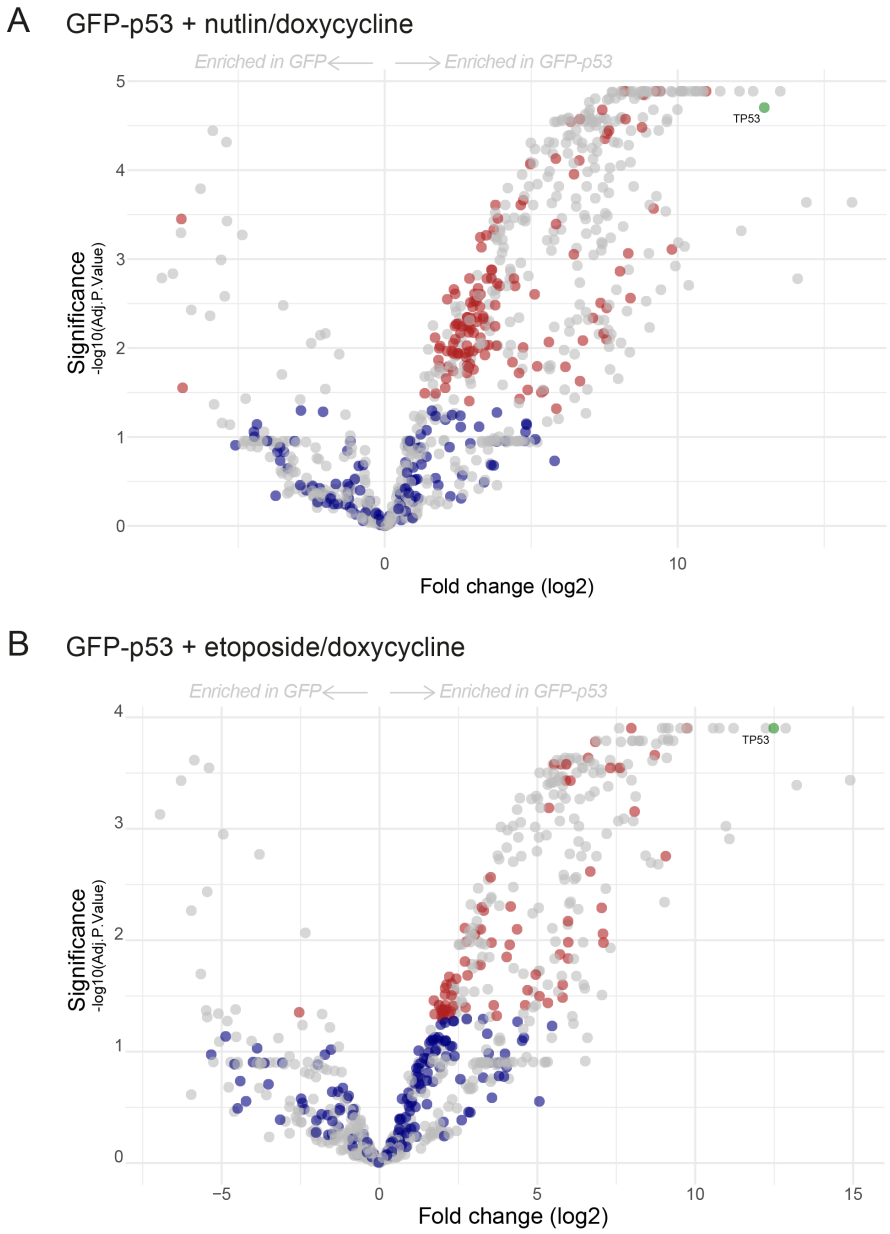


Figure 4.16: Proteomic analysis of GFP-p53 IPs reveals extensive binding to RBPs. Volcano plots for nutlin/doxycline GFP-p53 IPs (A) and etoposide/doxycycline GFP-p53 IPs (B), with RNA binding proteins (from Hentze *et al.* (2018)) highlighted. Red points denote RBPs with an adjusted P value < 0.05 in GFP-p53 versus GFP control. Blue points denote RBPs with an adjusted P value > 0.05 in GFP-p53 versus GFP control.

4.6 eGFP-p53 interacts with RNA in living cells

Once a suitable cellular system had been established, it was essential to confirm that eGFP-p53 retains the ability to interact with RNA. For this, I decided to use an approach complementary to RIC that could confirm the RNA-binding activity of p53 by an orthogonal approach. This approach, based on the initial stages of the irCLIP protocol [93], utilises UV crosslinking to covalently link protein-RNA interactions. This is followed by lysis, partial RNase digestion, and immunoprecipitation with GFP-trap. An L3 adaptor containing an infrared (IR) dye is then ligated onto the 3' end of the RNA. After SDS-PAGE and membrane transfer, the presence of protein-RNA complexes is determined by the presence of infrared signal at the expected molecular size. As the IR-conjugated adaptor is not commercially available, I generated it as described in the irCLIP (infrared CLIP) protocol [93] (Chapters 2 and 3). The L3 DNA adaptor with 5' phosphorylation, 3' biotinylation and an internal NHS-azide modification was pre-adenylated, before conjugation of the IR dye via click chemistry (Figure 4.17 A). A dot blot was then performed with serial dilutions of the resulting 1 uM IR linker stock, using in parallel 1 uM of the original IR linker kindly gifted by the Ule lab, which was used as a standard (Figure 4.17 B). Both L3-IR linkers exhibit similar fluorescence intensity, indicating the efficient conjugation of the IR dye to the L3 linker.

The IR linker was then used to label GFP-p53-RNA complexes. After UV-crosslinking, lysis, RNase treatment and immunoprecipitation using GFP-trap, the IR linker was then ligated onto the RNA 3' end, before visualisation of the IR dye using the LI-COR Odyssey Fc imaging system (Figure 4.18 A). Together with the nutlin/doxycycline treated GFP-p53 test sample, a non-induced GFP-p53 control

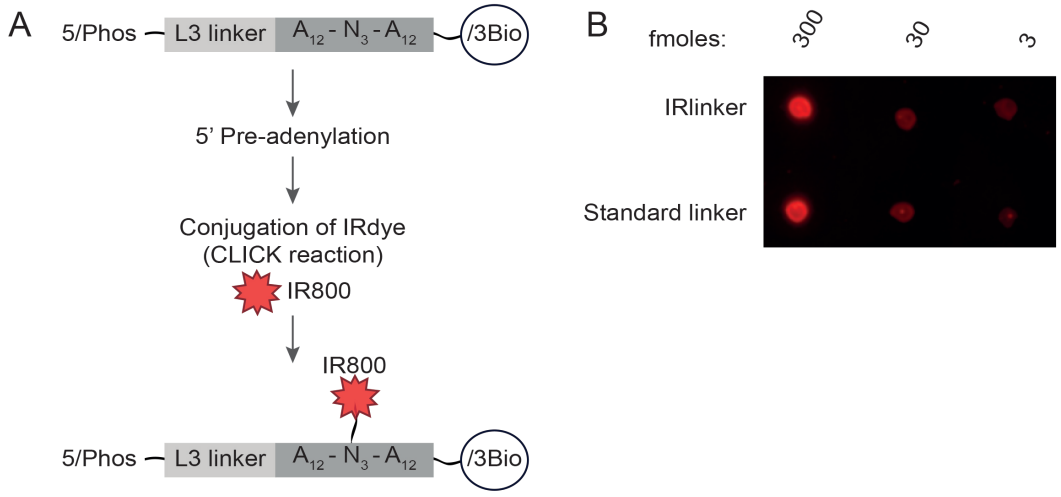


Figure 4.17: Production of IR linker for labelling of protein-RNA complexes. (A) Protocol schematic for production of IR linker. Briefly, a L3 DNA adaptor with 5' phosphorylation, 3' biotinylation and an internal NHS-azide modification was pre-adenylated, before conjugation of the IR dye via click chemistry. A dotblot (B) shows serial dilutions of generated IR linker, along with a standard IR linker (obtained from the Ule lab) to confirm successful conjugation of IR dye to the L3 DNA adaptor.

was included, as well as a nutlin/doxycycline treated NCL control (Figure 4.18 B). The IR linker contributes an additional 15 kDa onto the protein-RNA complex molecular weight. In this case, the IR linker plus GFP-p53 was predicted to be 90 kDa, so complexes including RNA fragments are expected to be larger than 90 kDa. We detected a clear smear over 90KDa that was dependent on the induction of GFP-p53 and UV crosslinking. This signal likely represents GFP-p53/RNA complexes, with the smear representing different molecular sizes due to heterogeneously digested RNA fragments generated by partial RNase digestion [94, 117, 95]. Non-induced and NCL controls show no signal at these predicted sizes, although a band is seen at approximately 90 kDa in the NCL control. The identity of this band is unclear. It may correspond to a truncated form of p53, which is supported by the presence of band with a slightly lower molecular weight than GFP-p53 in GFP-trap IPs (Figures

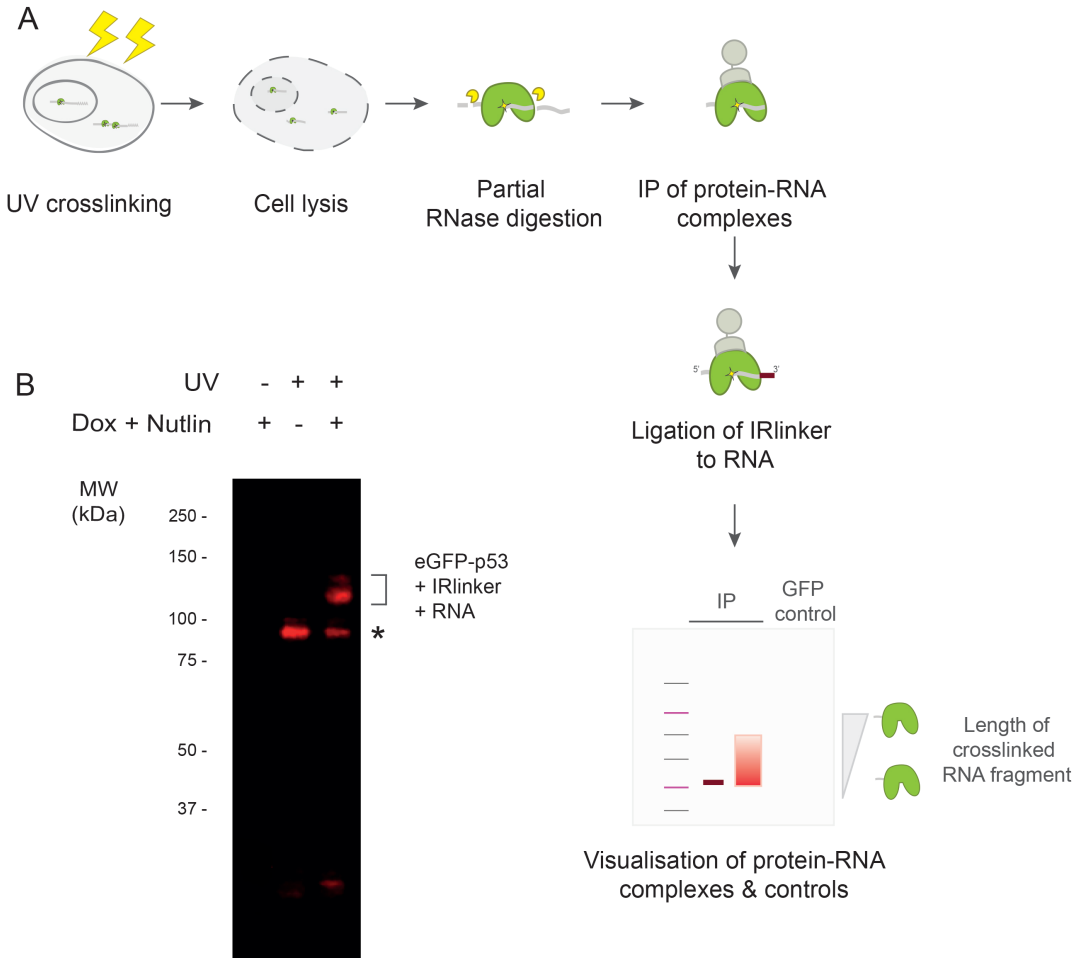


Figure 4.18: Visualisation of GFP-p53-RNA complexes. (A) Protocol for IR linker ligation to protein-RNA complexes. Cells are UV crosslinked and lysed. Partial RNA digestion is performed, followed by immunoprecipitation of RNA-protein complexes. The IR linker is then ligated to the RNA 3' end, followed by analysis by gel electrophoresis, transfer to nitrocellulose membrane and visualisation. (B) NCL GFP-p53, uninduced GFP-p53 and GFP-p53 treated with nutlin/doxycycline were processed according to the IR linker ligation protocol, and visualised. GFP-p53 + IR linker + RNA complexes are labelled. * may represent a smaller truncated GFP-p53, as can be seen in WB analysis, an endogenous p53 + IR linker + RNA, or a contaminant.

4.8 and 4.10). Alternatively, this band may correspond to GFP-p53 + linker.

4.7 Discussion

p53 has been shown to interact with RNA within acellular contexts; however, whether these interactions occur in living cells remained uncertain. In this chapter I have shown using a combination of orthogonal methods, that p53 is able to bind RNA in cellulo, using highly stringent conditions. In particular, I employed RNA centric (RIC) and protein-centric (irCLIP) approaches to verify the RNA-binding activity of p53 in living cells. Examination of p53's protein interaction network reveals extensive interaction with RBPs, suggesting p53's potential role as an RBP, and incorporation in multiple RNPs. These results extend previous (predominantly) *in vitro* studies highlighting p53's biochemical propensity to bind RNA, and work by Yoshida *et al.* [74] that showed potential *in vivo* interactions.

Previous attempts to characterise p53-RNA interactions *in vivo*, were limited mainly by the coimmunoprecipitation of contaminating RNPs using IgG antibodies [104], which made it ultimately impossible to distinguish between direct p53-RNA interactions and those caused by other unwanted RBPs. Indeed, in my initial experiments I also observed that antibodies typically used in p53 research (conventional IgG antibodies), and nanobodies, strongly contaminate the isolated p53 with unwanted proteins.

Our lab had already utilised the GFP-trap system, and recognised its increased specificity and affinity. As such, an obvious strategy was to generate a GFP-p53 expressing line. One of the potential limitations of such systems is the potential inability of GFP-p53 to interact with key protein partners, and indeed to interact

with RNA. Via proteomic analysis I was able to show GFP-p53 is still able to interact with expected p53 protein partners (Figures 4.14 and 4.15), and indeed oligomerise with endogenous p53 (Figures 4.8 A and 4.10 A). Finally, utilising an infrared linker combined with the initial stages of the irCLIP protocol (Figure 4.18 A), I was able to show that eGFP-p53 crosslinks to RNA in a UV-dependent manner (Figure 4.18 B), implying a bona fide interaction with RNA. My GFP-p53 system is thus suitable for follow up characterisation experiments, including CLIP-based methods, protein-protein interaction assays and functional assays, which will be approached in the forthcoming chapters.

5

Identifying the p53 RNA and protein interactomes

5.1 Introduction

Recent methodological and technological advances, such as RIC, have not only enabled identification of novel and non-canonical RBPs, but also further appreciation of the importance of RBPs in cellular function, homeostasis and disease. As such, efforts to understand the mechanisms, modalities and functions underlying these protein-RNA interactions have increased; resulting in significant advances in the ranges of techniques available to study protein-RNA interactions in greater detail.

The advent of crosslinking immunoprecipitation (CLIP) based approaches in the early 2000s [118] sparked a decade of progress in our ability to study protein-RNA interactions in living cells. These CLIP-based approaches, such as iCLIP [95], irCLIP [93], eCLIP [96] and spyCLIP [119], offer increased specificity and resolution, along with improved signal to noise ratios [120]. As the available experimental repertoire increases, proteins that were previously uncharacterised, poorly understood or difficult to study become viable candidates for experimental investigation. This chapter aims to utilise CLIP-based approaches, combined with the GFP-p53 cellular system described in Chapter 4, to decipher the role of p53 as an RBP. In particular, it is crucial to identify the network of RNAs recognised and bound by p53 in cellulo.

Traditionally, RBPs and their interactions with RNAs are considered from a protein-centric perspective. With this perspective, proteins interact with RNAs and exert a range of functions upon the RNA molecules, thus impacting their stability, localisation, translation and decay. Increasingly, however, the importance of RNA-centric outlooks is becoming more apparent. RNAs themselves can bind proteins, and impact their function, structure, localisation, and the formation of multi-protein complexes. In this chapter I will also investigate the potential roles of RNAs in

mediating p53-protein complex assembly.

Finally, I will use recent computational advances from our collaborators, to predict the p53 protein regions that are likely to promote its RNA binding activity.

5.2 Identifying the RNA targets of p53

To identify the scope of RNAs bound by p53 in living cells, a CLIP-seq based approach was used (henceforth termed CLIP-seq). Briefly, cells were UV irradiated to induce the formation of covalent crosslinks between proteins and RNA in direct contact. Cells were then lysed, followed by a stringent IP of the target protein. Proteinase K digestion of the eluates then enables liberation of bound RNAs, which are isolated, amplified via RT-PCR and sequenced (Figure 5.1). Unlike single nucleotide resolution CLIP methods [95], partial RNase digestion was not performed here, as my primary goal at this stage was to determine the network of RNAs interacting with p53, rather than specifying binding profiles or motifs. This protocol was based on the initial stages of the iCLIP protocol.

Maximising the efficiency of crosslinking of physiological protein-RNA complexes is crucial for enabling purification, and ultimately identification of, bound RNA species. Crosslinking efficiencies vary between proteins due to different binding geometries, interaction interface size, interaction half life, and sequence biases on either protein or RNA side [120, 121]. To determine the optimal conditions for GFP-p53 crosslinking, I used different UV doses, and measured the resulting RNA levels obtained after GFP-p53 IP, using chip-based capillary electrophoresis (Bioanalyser, Agilent). Crosslinking at 254 nm with 0.3 J/cm^2 resulted in a slightly higher RNA yield than 0.15 and 0.6 J/cm^2, and as such this dose was selected for subsequent

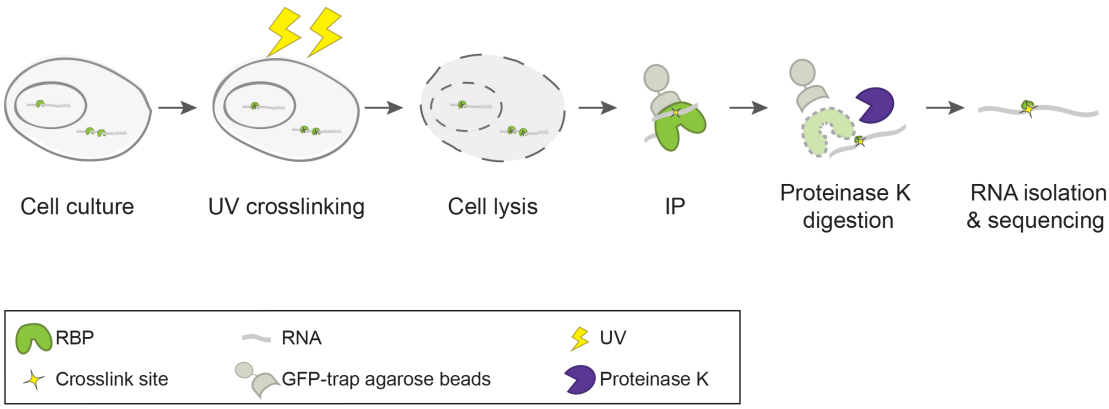


Figure 5.1: CLIP-seq protocol. Schematic representation of the CLIP-seq protocol. Cells are UV irradiated to induce the formation of covalent bonds between proteins and RNA in direct contact. Cells are then lysed, and protein-RNA complexes immunoprecipitated. Proteinase K digestion liberates bound RNA molecules, which can then be isolated. Sequencing libraries are produced using RT-PCR, and sequenced.

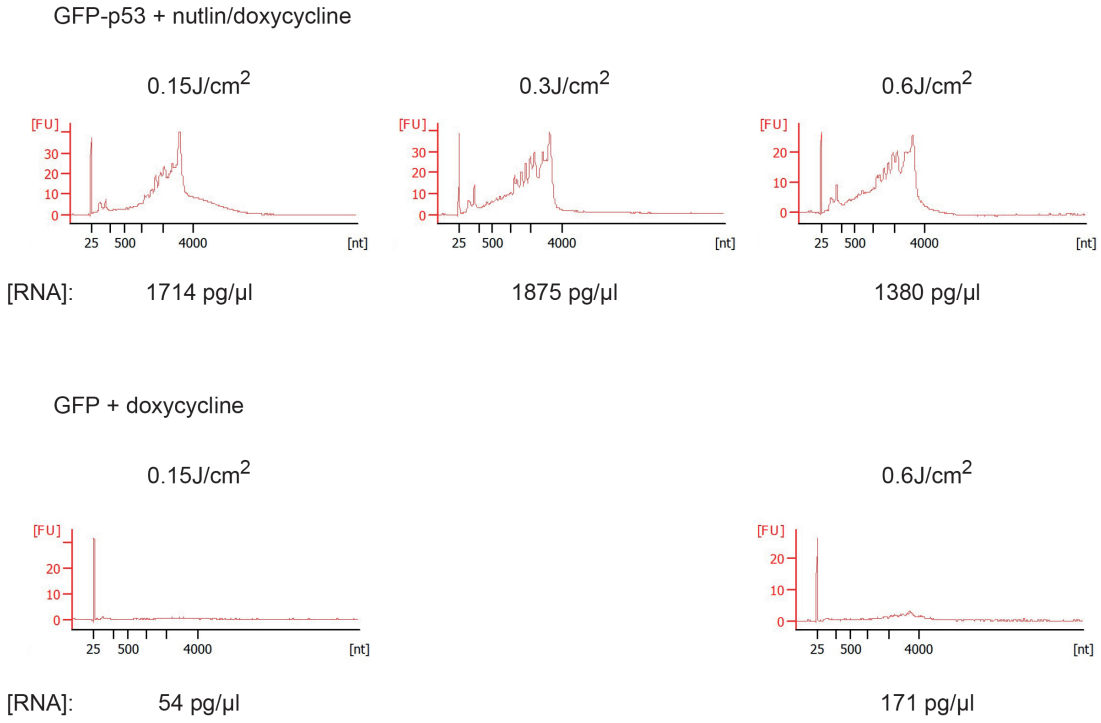


Figure 5.2: CLIP-seq UV crosslinking optimisation. UV crosslinking dosages were tested on GFP-p53 cells treated with nutlin/doxycycline. 245 nm UV was applied at 0.15, 0.3, and 0.6 J/cm^2. Control GFP cells treated with dox were also irradiated at 0.15 and 0.6J/cm^2. Shown are bioanalyser traces displaying RNA obtained from each proteinase K digested IP sample.

experiments (Figure 5.2).

GFP-p53 cells were treated with either nutlin and doxycycline or etoposide and doxycycline for 6 hours and subjected to the CLIP-seq protocol, across 3 replicates. It has been observed that the eGFP may co-precipitate highly abundant cellular RNAs [122]. Hence I followed the same protocol using a Hek293 Flp-In TREx cell line expressing eGFP as negative controls (in triplicate). Lysates' total GFP levels were measured in a platereader, and input material normalised to the sample with the lowest GFP reading. IPs were then performed, followed by liberation of protein-bound RNAs by proteinase K digestion. RNAs were quantified and converted into directional sequencing libraries using a commercially available dUTP based directional RNA sequencing library preparation kit, combined with single indices to enable sample multiplexing. Resulting libraries were mixed at equimolar ratios, and sequenced using single-read sequencing with single index reads.

Due to lack of a consensus CLIP-seq analysis pipeline, and the slight variations in the protocol used here, the CLIP-seq data was analysed according to a customised workflow, depicted in Figure 5.3. Demultiplexed reads were mapped to the human genome. Next, only reads that uniquely mapped to the genome were included, and of these only those that did not span splice junctions. Subsequent analytical steps were performed on these uniquely mapping, non-spliced reads. GFP-p53 samples were then compared to their respective eGFP controls, to remove contaminant RNAs non-specifically bound to the eGFP tag. Genomic bins where enrichment of the GFP-p53:GFP ratio was less than 4 fold (Log2>2) were discarded. The number of reads mapping to genomic features were then counted.

Previous works initially highlighted p53's ability to bind to protein coding mRNAs , along with more recently ncRNAs [83, 85]. To determine which types of RNAs

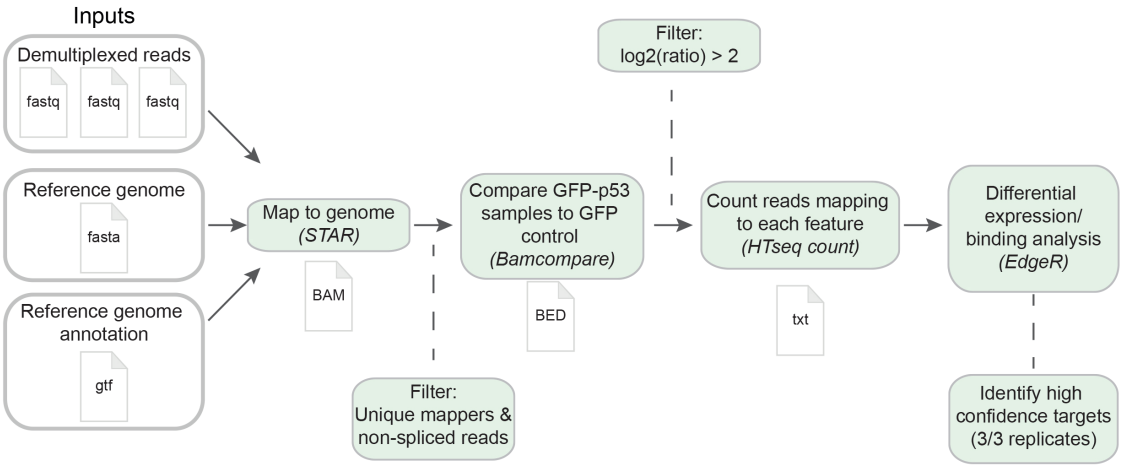


Figure 5.3: CLIP-seq analysis pipeline. Schematic showing the analysis pipeline developed for CLIP-seq samples. Demultiplexed fastq files were aligned to the human genome using the STAR RNA-seq aligner. Multimapping and spliced reads were removed. GFP-p53 sample signal was then compared to GFP control signal across the genome, and only regions with log2(ratio GFP-p53 over GFP) $>$ 2 were retrieved. Genomic features and biotypes were then counted using HTseq count. Edge R was then used to perform differential expression analysis on resulting read counts. Additionally, high confidence targets were identified by assessing presence of transcripts in all 3 replicates for each GFP-p53 treatment condition.

are bound by p53, we classified those RNAs identified by CLIP-seq that passed the enrichment-over-control threshold, based on biotype annotation. Importantly, 75-85% of eGFP-p53 reads map to protein coding, messenger (m)RNAs, with approximately 10-15% mitochondrial rRNA, and the rest a combination of pseudogenes, antisense transcripts and lincRNAs (not shown). After removal of multimapping and spliced reads, over 90% of reads correspond to protein coding transcripts, with effective removal of mitochondrial rRNAs (Figure 5.4). Given that protein-coding transcripts only account for about 3-7% of the mammalian transcriptome [7], this enrichment strongly indicates that p53 interacts predominantly with mRNAs. Reassuringly, this proportion is similar to those previously observed in eCLIP and iCLIP data for bona-fide mRNA-binding proteins (mRBPs). These data strongly agree with *in vitro* experiments in which p53 was proposed to interact with mRNAs.

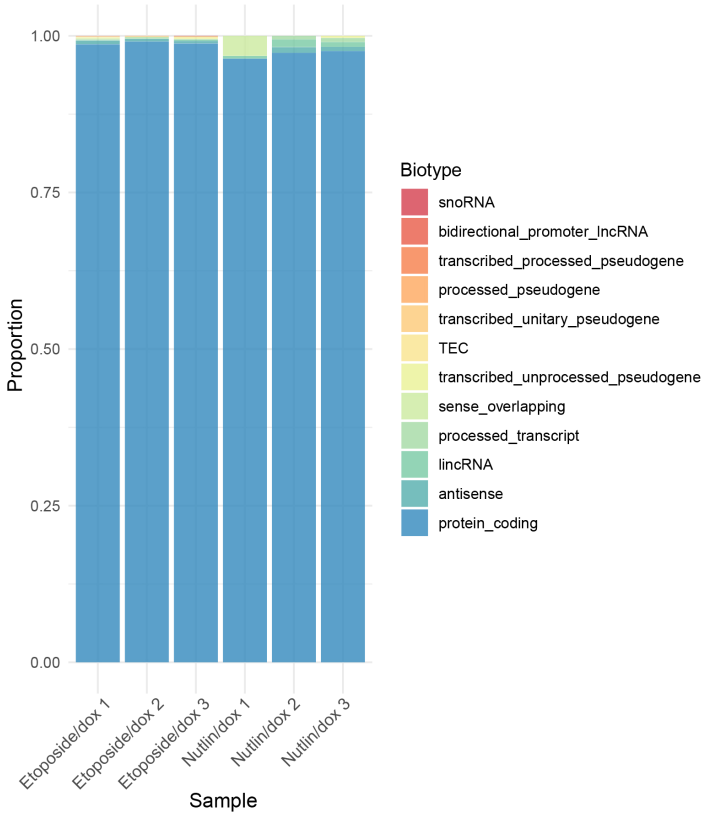


Figure 5.4: p53 binds mainly protein coding transcripts. After all processing steps have been completed, types of transcripts bound (biotype) in each sample were counted, and plotted as a proportion of the whole sample.

Interestingly, these *in vitro* studies proposed that p53 could interact with the mRNAs encoding its co-factors *MDM2* and *MDM4*, downstream p53 response genes *CDKN1A, CDK4*, and *FGF-2*, and *p53* itself, invoking potential post-transcriptional regulatory feedback loops, and the ability of p53 to regulate the expression of core target genes post-transcriptionally, as well as via its transcription factor function. To determine if these *in vitro* results are recapitulated in cellulo, I interrogated my CLIP-seq datasets (Table 5.1). Notably, both *p53* and *MDM2* mRNAs are highly enriched in all three replicates of each condition (i.e with nutlin or etoposide), suggesting that such a feedback loop may exist in cells. However, *CDKN1A* and *MDM4* mRNAs were not identified in all replicates or conditions, respectively. Finally, *CDK4* was only identified in 1 replicate of etoposide-treated cells, while *FGF-2* was detected in none. Hence, *p53, MDM2* and *MDM4* mRNAs emerge here as bona fide targets of p53 in a treatment-dependent context, the latter only in etoposide-treated cells. Whether p53 interacts with *CDKN1A* mRNAs is debatable, whilst this data does not support p53 binding to *CDK-4* and *FGF-2* mRNAs. Interestingly, the fact that *MDM4* binding to p53 is only detected in cells undergoing a DNA damage response (etoposide), raises the notion that p53's RNA binding may be varied depending upon the specific stress stimuli, as is seen for its DNA binding activity [123]. In conclusion, this dataset validates previously identified p53 bound RNA candidates, including *p53, MDM2*, and, possibly, *CDKN1A*, whilst highlighting the potential for stimulus-specific candidates, such as *MDM4*.

Table 5.1: Validating p53 *in vitro* interactors.

Gene Name	**Nutlin/dox replicates**	**Etoposide/dox replicates**
TP53	3	3

Continuation of Table 5.1

Gene Name	Nutlin/dox replicates	Etoposide/dox replicates
MDM2	3	3
MDM4	0	3
FGF-2	0	0
CDKN1A	2	2
CDK4	0	1

Recent works in the p53 field have focussed on defining p53's core transcriptional programme – that is, genes that are consistently identified as regulated by p53's transcription factor function, across a range of cell types, stressors and cellular responses [123]. In a similar vein, defining p53's core RNA targets, shared across a range of p53 activating conditions, would be informative, and provide insight into the preliminary functions of p53's interactions with RNA. The top 10 p53 bound transcripts are shown for both nutlin treated (Table 5.2) and etoposide treated (Table 5.3) cells.

Table 5.2: GFP-p53 + Nutlin/Doxycycline top candidate RNA targets. Candidates were selected based on a log2(fold change) > 2 in GFP-p53 samples compared to GFP controls. Candidates had to be identified in all 3 GFP-p53 replicates.

Gene Name	Total reads across replicates
TP53	3447
CBX6	275
CAPN1	217
SBF1	134
MICALL1	78

Continuation of Table 5.2

Gene Name	Total reads across replicates
MDM2	68
CEP170B	59
NAT8L	56
SPEN	53
SPHANK3	5

Table 5.3: GFP-p53 + Etoposide/Doxycycline top candidate RNA targets. Candidates were selected based on a log2(fold change) > 2 in GFP-p53 samples compared to GFP controls. Candidates had to be identified in all 3 GFP-p53 replicates.

Gene Name	Total reads across replicates
TP53	4885
PRRC2B	3260
JUN	2398
KMT2D	1819
EGR1	1148
PRRC2A	834
TCOF1	747
MICALL1	556
UBR4	470
URB1	469

These high confidence gene sets (log2(fold change) greater than 2 in GFP-p53 over GFP, present in all 3 samples) were then compared, and a group of 33 shared

candidates identified, that were present in all 3 replicates in both treatment conditions (Figure 5.5, Table 5.4).

Table 5.4: Shared RNA targets. Candidates selected for each of the induction conditions (nutlin/doxycycline or etoposide/doxycycline) were compared. Only those present in all 3 replicates for both induction conditions were selected.

Gene name	Gene name
TP53	*XPO5*
KMT2D	*PPRC1*
PRRC2A	*SPEN*
MICALL1	*NUP153*
CBX6	*MYBBP1A*
RRP7A	*MPRIP*
CEP170B	*CHST3*
NF2	*FOXK1*
SEC16A	*CUL7*
SBF1	*SHANK3*
TRRAP	*ANKRD11*
MDM2	*SRF*
FBRSL1	*INTS1*
MYH9	*NUP98*
KDM5C	*PHIP*
DBNL	*EWSR1*
DPY19L3	

Notably, of the 33 shared RNAs identified, 17 are linked to p53 in the literature

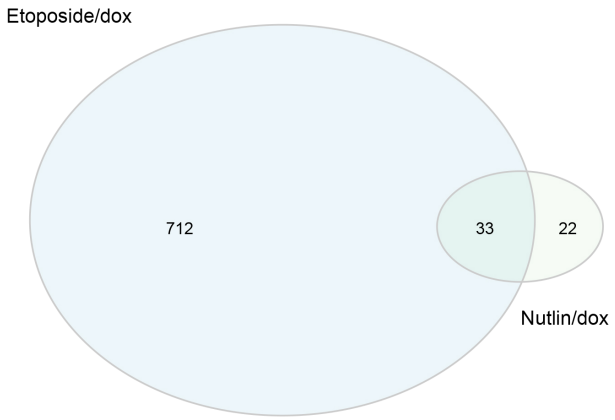


Figure 5.5: High confidence GFP-p53 bound transcripts. Transcripts bound in 3 out of 3 replicates for etoposide/dox and nutlin/dox were retrieved. 745 etoposide/dox bound transcripts were identified, and 55 nutlin/dox transcripts, with an overlap of 33 shared high confidence interactors.

(Table 5.4). *MDM2, MICALL1* and *p53* itself are known 'core' p53 transcriptional targets [123]. Another RNA target with a well-defined association with p53 is *CUL7*, which is an E3 ubiquitin ligase and member of the 3M and CUL7-RING complexes, that ubiquitinates p53 along with MDM2. All these, together with the presence of the p53-inhibitor *MDM4* upon etoposide treatment, suggest plausible post-transcriptional autoregulatory feedback loop, involving the regulation of p53 itself, and several of its negative regulators.

Moreover, p53 targets Myb-binding protein 1a (*MYBBP1A*) mRNA. MYBBP1A translocates from the nucleolus to the nucleoplasm upon nucleolar stress, whereupon it facilitates p53 oligomerisation and acetylation [124, 125]. This represents an additional regulator of p53, whose cognate RNA is a target of p53. Additional mRNA targets include histone modifiers (*KMT2D* and *KDM5C*) and histone acetyltransferase complex component *TRRAP*. Additionally, two nucleoporins (*NUP98* and *NUP153*) are identified as high confidence candidates, along with *XPO5*, all of

which are involved in nucleocytoplasmic transport of macromolecules.

To determine whether the p53 activating stimulus impacts p53 RNA binding activity or preferences, we compared the CLIP-seq datasets generated for etoposide or nutlin treated cells. Principle component analysis (PCA - Figure 5.6 A) and hierarchical clustering (Figure 5.6 B) show that replicates of each experimental condition share more similarities than when compared to replicates from the other experimental condition. One of the IP replicates of etoposide-treated cells shows some degree of dissimilarity in PCA (Figure 5.6 A). However, it still clustered together with the other etoposide treated replicates, and shared a number of the most abundant RNAs (Figure 5.6 B).

The volcano plot in Figure 5.7 shows in colour the transcripts differentially co-IPed with p53 in etoposide-treated versus nutlin-treated cells, along with a larger group of transcripts in grey that do not exhibit significant differences between the two experimental conditions. Both groups of transcripts can provide us with important information.

Grey points on the volcano represent RNAs equally bound by p53 in the two treatment conditions, and therefore represent the 'core' target RNA network of p53. This group accounts for 26 of the high-confidence shared p53 interactors highlighted above. However, 150 additional transcripts are identified in this group, falling under a lower confidence threshold and thus representing potential candidate RNAs. The lower significance of these RNAs is generally due to their presence in only 2 of the 3 replicates, particularly in the nutlin-treated samples. As the read count is the limiting factor in IPs from nutlin-treated samples, I considered any of these non-significant hits with greater than or equal to 30 reads in nutlin treated samples,

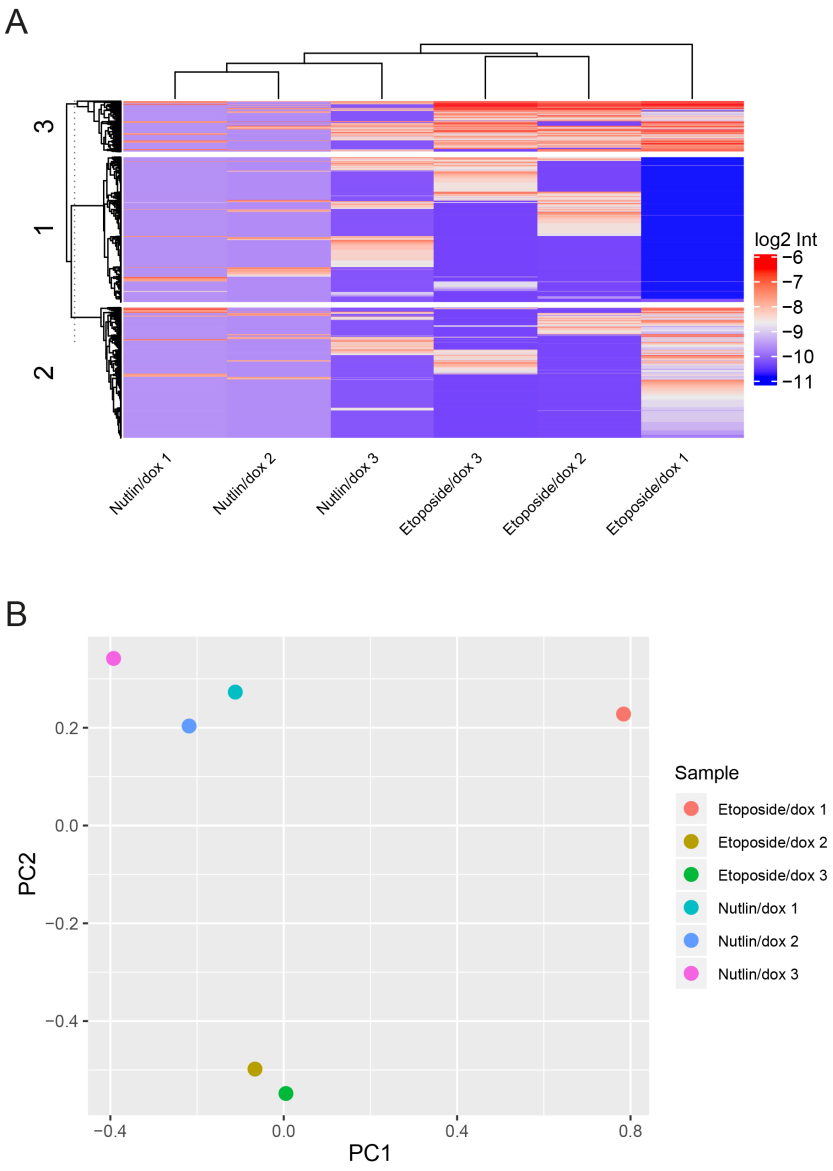


Figure 5.6: PCA and hierarchical clustering of GFP-p53 bound targets. Hierarchical heat map clustering (A) and principal components analysis (B) was performed on normalised pseudocounts of all transcripts enriched over GFP controls.

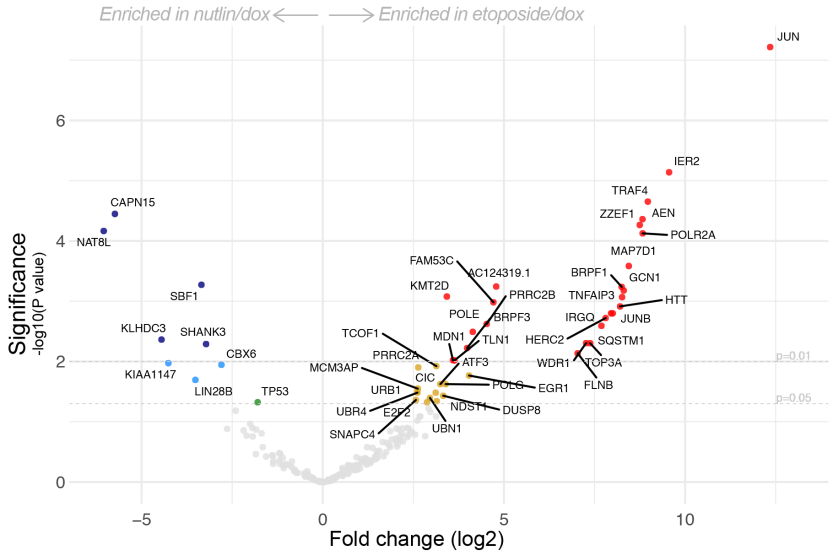


Figure 5.7: Differential binding based on GFP-p53 induction condition. Volcano plot shows the differential binding of transcripts in etoposide/dox treated GFP-p53 cells compared to nutlin/dox treated GFP-p53 cells. Log2 fold change is plotted against -log10 adjusted P value. Hits with a P value < 0.05 are highlighted in light blue and yellow, and those with p-value < 0.01 are highlighted in dark blue and red. Transcripts that do not reach either significance threshold are denoted in grey.

and present in at least 2 out of 3 replicates in both treatment conditions. This highlighted an additional 5 transcripts as 'candidates' – eukaryotic initiation factor 5 (*eIF5*), TBC1 Domain Family Member 16 (*TBC1D16*), LDL Receptor Related Protein 3 (*LRP3*), Family with sequence similarity 53 Member B (*FAM53B*) and finally, N-Myc Downstream Regulated 1 (*NDRG1*). Differential analysis therefore reveals 38 candidate p53-bound RNAs, including 33 high confidence candidates. Of these additional candidates, NDRG1 is a metastasis suppressor, which has been associated with p53 in several studies, either as a p53 transcriptional target [126], or as a regulator of p53 pathway genes, such as p21 [127]. NDRG1 has been shown to be required for p53-mediated caspase activation, apoptosis [128] and mitotic spindle checkpoint [129]. The exact molecular functions of NDRG1 remain unclear. Interestingly, NDRG1, along with NDRG2 and NDRG4 were recently identified as

harbouring RNA binding activity themselves in RBDmap datasets [15]. 10 of the 38 shared candidate transcripts are previously characterised RBPs [11], with GO "RNA binding" annotation also present for p53 and MDM2.

In contrast, points in figure 5.7 highlighted in colour represent transcripts that are differentially enriched in the IPs performed with the two different treatments. 158 transcripts are more abundant in IPs performed from etoposide-treated cells, which represent potential DNA damage specific target transcripts. Immunoprecipitaton of p53 in etoposide treated samples did not only result in a larger pool of co-isolated RNAs compared to nutlin treated samples, but also to substantially higher read counts per transcript in general terms. This appears to be a p53 induction-specific phenomenon, derived from the intrinsic differences between nutlin (nongenotoxic) and etoposide (genotoxic) induction conditions.

To summarise, CLIP-seq reveals a group of 38 candidates, which represent core p53 RNA targets, present under both genotoxic and nongenotoxic p53 activating conditions. A number of these candidates are known p53 targets, and regulators, suggesting the presence of post-transcriptional autoregulatory feedback loops, similar to those observed at the transcriptional level. Further work is required to elucidate the function of these p53-RNA interactions. Large differences in the number of RNAs bound by p53 under the two experimental conditions used likely reflects stimulus-specific effects, with DNA damage appearing to induce higher levels of p53-RNA interactions.

5.3 Understanding p53-RNA binding at single nucleotide resolution

Utilisation of a CLIP-seq approach has enabled the identification of potential p53 RNA interaction partners. However, lack of RNase digestion within this protocol prevents resolution of the RNA regions recognised by p53. Therefore, whilst we have been able to elucidate some of the transcripts bound by p53, still lacking is information regarding the modality of these interactions. Without this binding site information, we cannot elucidate whether p53 recognises sequence motifs and/or structural elements, or whether binding is nonspecific, potentially driven indirectly through p53's extensive protein-protein interactions. To circumvent this limitation, we applied an individual nucleotide resolution UV crosslinking immunoprecipitation (iCLIP [95]) protocol to GFP-p53 cells treated with either nutlin or etoposide (in combination with doxycycline).

iCLIP enables the identification of protein-RNA binding sites to single nucleotide resolution, by taking advantage of the propensity of reverse transcriptase enzymes to terminate reverse transcription (RT) when encountering peptide fragment remnants at UV crosslink sites [95]. By examining the RT termination sites, we can pinpoint the exact nucleotide position at which a protein crosslinks to RNA (Figure 5.8). Briefly, (see Figure 5.8) iCLIP utilises UV irradiation to induce the formation of covalent crosslinks between proteins and RNA in direct contact. Cells are lysed, and partial RNase digestion performed to enable purification of RNA fragments of a desired size. Protein-RNA complexes are then purified under stringent conditions. To overcome lack of access to radioactive isotopes in our laboratory, I utilised the

infrared linker described previously. RNA is dephosphorylated, and a linker with an infrared fluorophore is ligated to the 3' end of the RNA fragment. This enables visualisation of protein-RNA complexes after SDS-PAGE and membrane transfer, by virtue of the IRlinker's infrared dye. Complexes can then be size selected from the membrane, allowing specific extraction of target protein-RNA complexes. Proteinase K digestion is performed, and the resulting RNA extracted. iCLIP then utilises specific RT primers, comprising of a cleavable adaptor region, Illumina sequencing adaptor sequences, as well as experimental barcodes (to allow demultiplexing after sequencing) and unique identifier sequences (to allow removal of PCR amplification products). Resulting cDNA is then size selected and purified, before the cDNA molecule is circularised. This rearrangement of the cDNA results in the crosslinked site (and site of RT termination in 80% of cases [120]) to be the first nucleotide after the sequencing adaptor. cDNA can then be amplified by PCR, size selected and sequenced. iCLIP is a long, technically demanding protocol. I helped establish the iCLIP protocol in our lab early in my PhD. Subsequent application of iCLIP to a new protein requires extensive optimisation of crosslinking, immunoprecipitation, RNase digestion and cDNA amplification conditions [120].

The use of appropriate positive and negative controls is integral for optimisation of iCLIP conditions [120, 117]. As such, initial stages of iCLIP optimisation were performed using GFP-p53 immunoprecipitates from nutlin/dox-treated cells, along with positive controls (hnRNPC-GFP), and a range of negative controls (GFP, non-induced GFP-p53 and nonirradiated samples, termed NCL).

As iCLIP is reliant on highly specific purification of protein-RNA complexes, I first assessed the specificity of the GFP-trap IP using iCLIP lysis and wash condi-

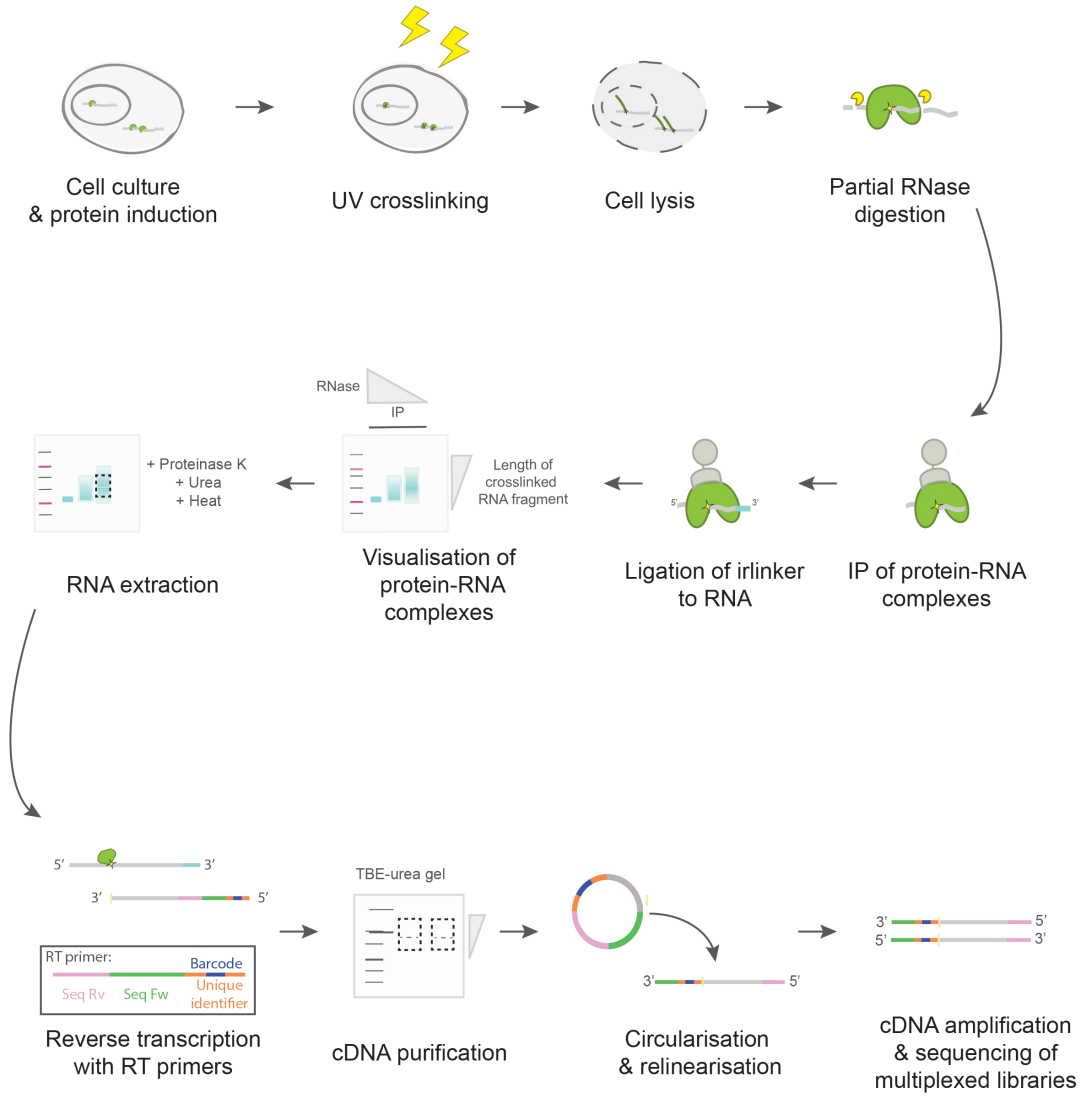


Figure 5.8: iCLIP protocol. Schematic representation of the iCLIP protocol. Cells are UV irradiated to induce the formation of covalent bonds between proteins and RNA in direct contact. Cells are then lysed, and partial RNase digestion performed. Protein-RNA complexes are then immunoprecipitated. An infrared dye-conjugated DNA linker is then ligated onto the RNA's 3' end, which enables visualisation of protein-RNA complexes after size seperation by SDS-PAGE and membrane transfer. RNA can then be size selected and extracted from the membrane via proteinase K digestion accompanied by 7 M urea. RNA extraction is followed by conversion of the RNA into cDNA via reverse transcription, using special RT primers. These contain a cleavable adaptor region, unique experimental barcode to enable demultiplexing, and additional random barcode sequences to facilitate removal of PCR duplicates. cDNA is then purified and size selected via gel electrophoresis using a TBE-urea gel. The resulting cDNA is circularised, and relinearised using the RT primer's cleavable region. Resulting cDNA is amplified via PCR. Multiplexed libraries are then generated and sequenced.

tions. Samples were UV irradiated, lysed, and partial RNase digestion performed using a 1:50 dilution of RNase I. Protein-RNA complexes were immunoprecipitated and IRlinker ligation performed. Samples were then subject to SDS-PAGE and either silver staining (Figure 5.9 A) or western blotting (Figure 5.9 B, C and D). Silver staining of the selected IPs across 2 replicates demonstrates that the target proteins represent the dominant bands in each sample, with a band of lower molecular weight, corresponding to the RNase enzyme used for partial RNA digestion. Importantly, there are no bands corresponding to protein species of higher molecular weight than the target protein, which is key as higher molecular weight proteins, and their bound RNAs, can contaminate RNA preparation during the size-selection step. Immunoblot with an anti-GFP antibody shows significant enrichment of target proteins (Figure 5.9 B, C and D) compared to inputs. UV protein-RNA crosslinking together with the high specificity and affinity of the eGFP nanotrap, allows the use of high-salt and ionic detergents during cell lysis and IP washes, which in turn ensures disruption of protein-protein and non-covalently linked protein-RNA interactions. This not only enables removal of non-target proteins and contaminating RNAs, but, importantly, ensures sufficient antibody accessibility to protein-RNA complexes [120].

An additional crucial stage of the protocol that requires optimisation is the partial RNase digestion performed during lysate preparation. Optimal RNase digestion is required to avoid copurification of multiple RBPs that crosslink to the same RNA molecule, whilst also ensuring sufficiently sized cDNAs within the final sequencing library to enable mapping to repetitive genomic regions [120]. Optimal size range for iCLIP RNA fragments after partial RNase digestion is 30-200 nucleotides

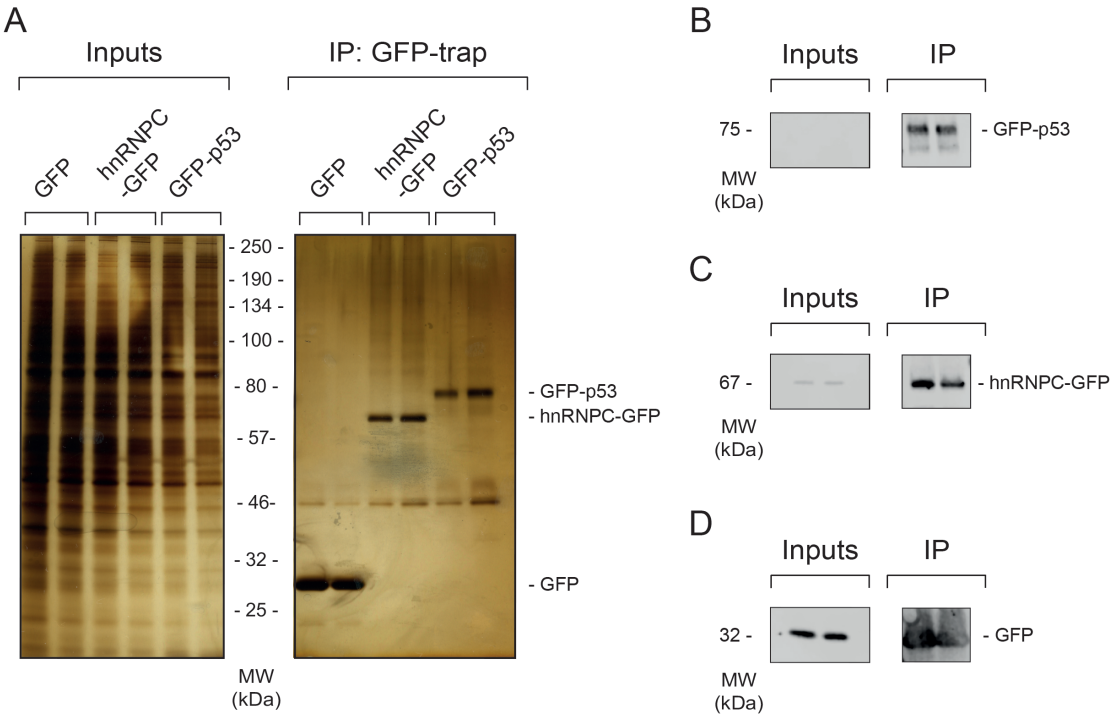


Figure 5.9: Testing iCLIP immunoprecipitation conditions. The iCLIP protocol was performed on GFP and hnRNPC-GFP control samples, along with a nutlin/dox induced GFP-p53 sample, up until size seperation of protein-RNA complexes by SDS-PAGE. Samples were then analysed by silver staining (A) and western blotting against GFP for GFP-p53 (B), hnRNPC-GFP (C) and GFP only (D).

[120]. GFP-p53 immunoprecipitates were subject to partial RNase digestion using 1:50, 1:100 and 1:250 dilutions of RNase I. Samples were subsequently digested with proteinase K, followed by RNA isolation. At this step, RNA was examined using chip-based capillary electrophoresis (Bioanalyser, Agilent). The resulting traces are shown in Figure 5.10. RNase digestion using the 1:250 RNase I dilution gave the best RNA yield, in terms of RNA concentration and, more importantly, size, with RNA fragments peaking around 100 nts (Figure 5.10 C). It is worth noting that even though co-purified RNA is optimal in terms of size distribution, the yield is limited. This is likely due to the loss of material linked to the low efficiency of the membrane transfer and proteinase K release step, as well as the RNA precipitation procedures. Hence, I decided to double the amount of starting material (2x 15 cm dishes, approximately 80% confluent upon induction).

During optimisation of the CLIP-seq protocol, I concluded that UV irradiation using a UV dosage of 0.3 J/cm^2 at 254 nm wavelength was optimal for the GFP-p53 protein. Subsequently, I wanted to confirm this was the case when performing the iCLIP protocol. GFP-p53 immunoprecipitates from cells irradiated with either 0.15 J/cm^2 or 0.3 J/cm^2 UV dosages were subject to the full iCLIP protocol, using 1:250 RNase I for RNA fragmentation. A positive control hnRNPC-GFP irradiated with 0.15 J/cm^2 was also included. Test PCRs were performed on obtained cDNA using 20, 25 and 30 cycles, and resulting amplification adjudicated by analysis of cDNA on a TBE gel, combined with DNA staining (Figure 5.11). The control hnRNPC-GFP shows sufficient cDNA amplification by 25 cycles, and potential overamplification by 30 cycles. Contrastingly, GFP-p53 samples only show discernable amplification after 30 cycles, and only when UV irradiation is performed using 0.3 J/cm^2 dosage.

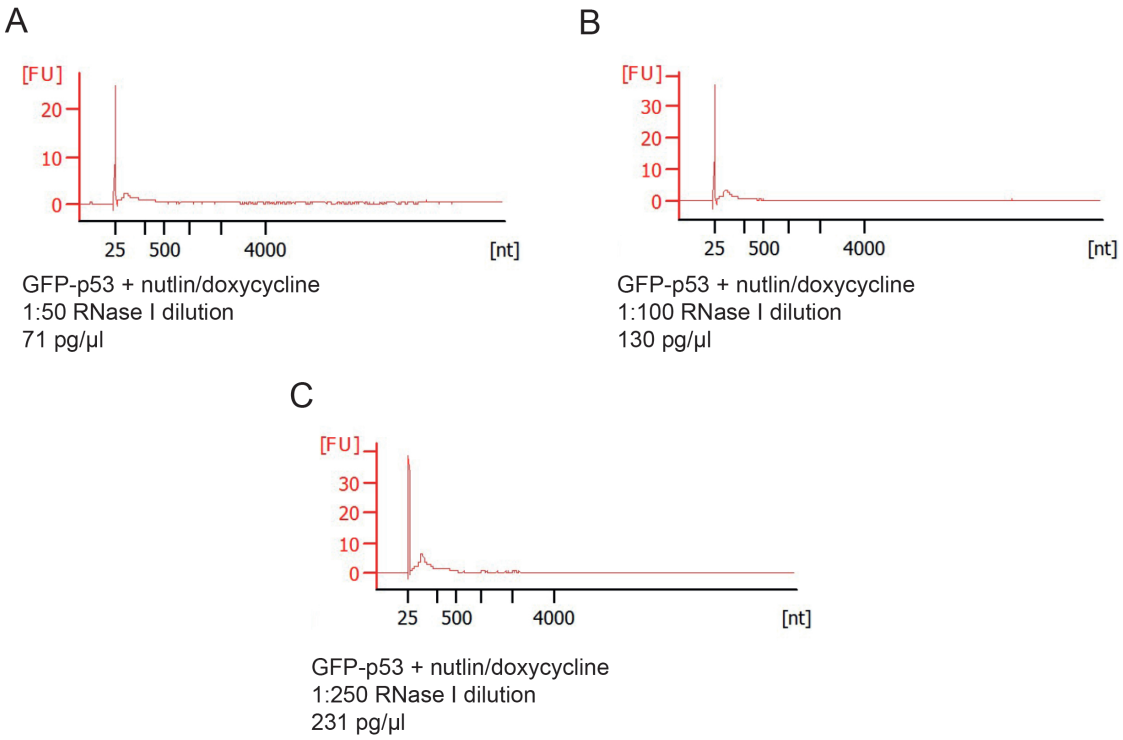


Figure 5.10: Optimising iCLIP partial RNase digestion. The iCLIP protocol was performed using GFP-p53 samples treated with nutlin/doxycycline, and crosslinked with 0.3 J/cm^2 UV dosage at 254 nm. Partial RNase digestion was performed using RNase I dilutions of 1:50 (A), 1:100 (B), and 1:250 (C). GFP-trap IP was performed, protein-RNA complexes seperated by SDS-PAGE followed by membrane transfer. Protein-RNA complexes were size selected, and proteinase K digestion performed. RNA was then isolated and analysed by chip-based capillary electrophoresis using the Agilent bioanalyser, mRNA pico kit. Fluorescence units (FU) are displayed on the Y axis, with size in nucleotides on the X-axis.

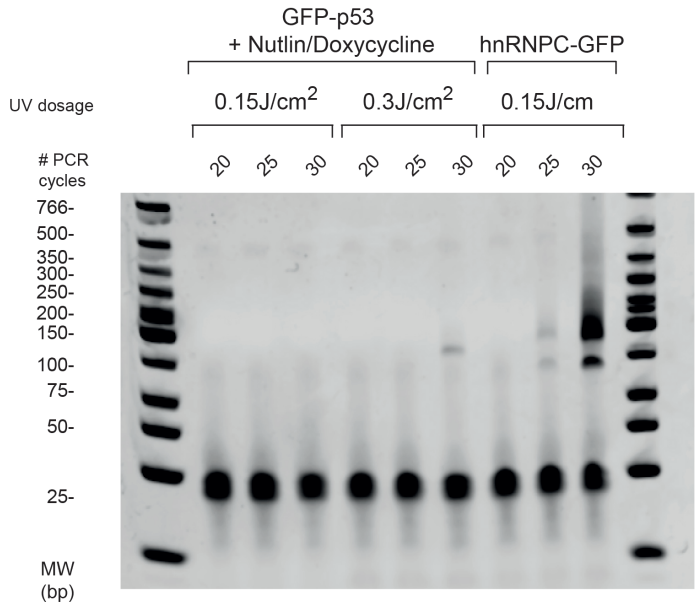


Figure 5.11: Optimising iCLIP UV crosslinking conditions. The iCLIP protocol was performed using GFP-p53 samples treated with nutlin/doxycycline, and crosslinked with either 0.15 or 0.3 J/cm^2 UV dosages at 254 nm. A positive control was included (hnRNPC-GFP + dox, irradiate with 0.15 J/cm^2 254 nm). cDNA amplification was performed using a range of cycles (20,25,30), and the resulting cDNA visualised on a TBE gel, stained with Sybrgold, to assess presence of successfully amplified cDNA. Lower bands represent primers, and higher bands represent amplified cDNA.

As such, all subsequent samples, including GFP controls, were irradiated using this UV dosage.

Once IP conditions, RNase digestion and UV crosslinking conditions were all established, the full iCLIP protocol was performed, and cDNA amplification conditions optimised. To control for potential non-specific binding of RNAs to the eGFP tag, paired GFP controls were included, treated with the corresponding etoposide or nutlin, in combination with doxycycline. Corresponding GFP-p53 samples were generated, along with a hnRNPC-GFP positive control. Test PCRs were performed to establish correct PCR amplification conditions for the final libraries (Figure 5.12 A and B, hnRNPC-GFP not shown), with a no DNA PCR control included. The optimal size of final PCR products ranges from 145-225 nucleotides [94]. The small-

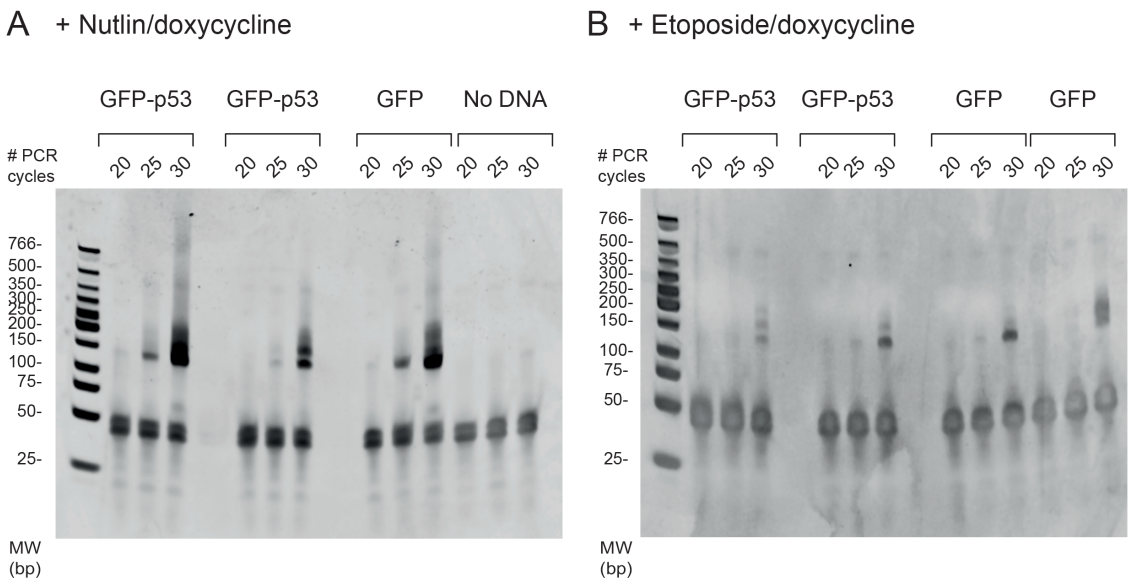


Figure 5.12: iCLIP test PCRs. The iCLIP protocol was performed on GFP-p53 and GFP samples treated with nutlin/dox (A) and etoposide/dox (B). Test PCRs were performed to assess the optimal number of PCR cycles for cDNA amplification for sequencing, and the resulting cDNA visualised on a TBE gel, stained with Sybrgold.

est band observed (Figure 5.12 A and B) across the samples is at the lower range of this size, and may be more susceptible to PCR artefacts [117, 120] than bands of a larger size, which can be seen with increased number of PCR cycles. Additional test PCRs were performed to fine-tune the final PCR cycling conditions (not shown). Final PCRs used the minimum number of PCR cycles required for a cDNA band of the correct size to be observed, less 1 cycle, and are denoted in Table 5.5. Two sets of final PCRs were performed using the remaining samples, split in half, and the resulting cDNA quantified, and combined. Multiplexed libraries were generated, using equimolar amounts of each sample. The library was then sequenced using single end sequencing.

Table 5.5: iCLIP - final library PCR cycling conditions.

Sample	Number PCR cycles
GFP-p53 + nutlin/doxycycline 1	25
GFP-p53 + nutlin/doxycycline 2	24
GFP-p53 + nutlin/doxycycline 3	25
GFP + nutlin/doxycycline	25
GFP-p53 + etoposide/doxycycline 1	27
GFP-p53 + etoposide/doxycycline 2	27
GFP + etoposide/doxycycline 1	26
GFP + etoposide/doxycycline 2	25

iCLIP results represent unique challenges in terms of downstream analysis. The iCount analysis pipeline (Curk *et al.*, in preparation) has been developed in order to address these challenges (Figure 5.13). It enables demultiplexing and subsequent alignment of reads to a generated genome index. The first nucleotide of each sequence represents the nucleotide downstream of the protein-RNA crosslink sites. These crosslink sites can be quantified, and categorised based on whether they represent uniquely or multiply mapped sites. Individual cDNAs or reads can be used in this case, which enables estimation of the level of overamplification that has occurred within the experiment. Individual crosslink events are then grouped into peaks, and significant peaks identified. Peaks are then clustered into significant clusters. The iCount pipeline then facilitates annotation of each site based on numerous characteristics, such as subtype of transcript bound. Summary statistics can also be provided for each significant site. I will implement a modification of the protocol,

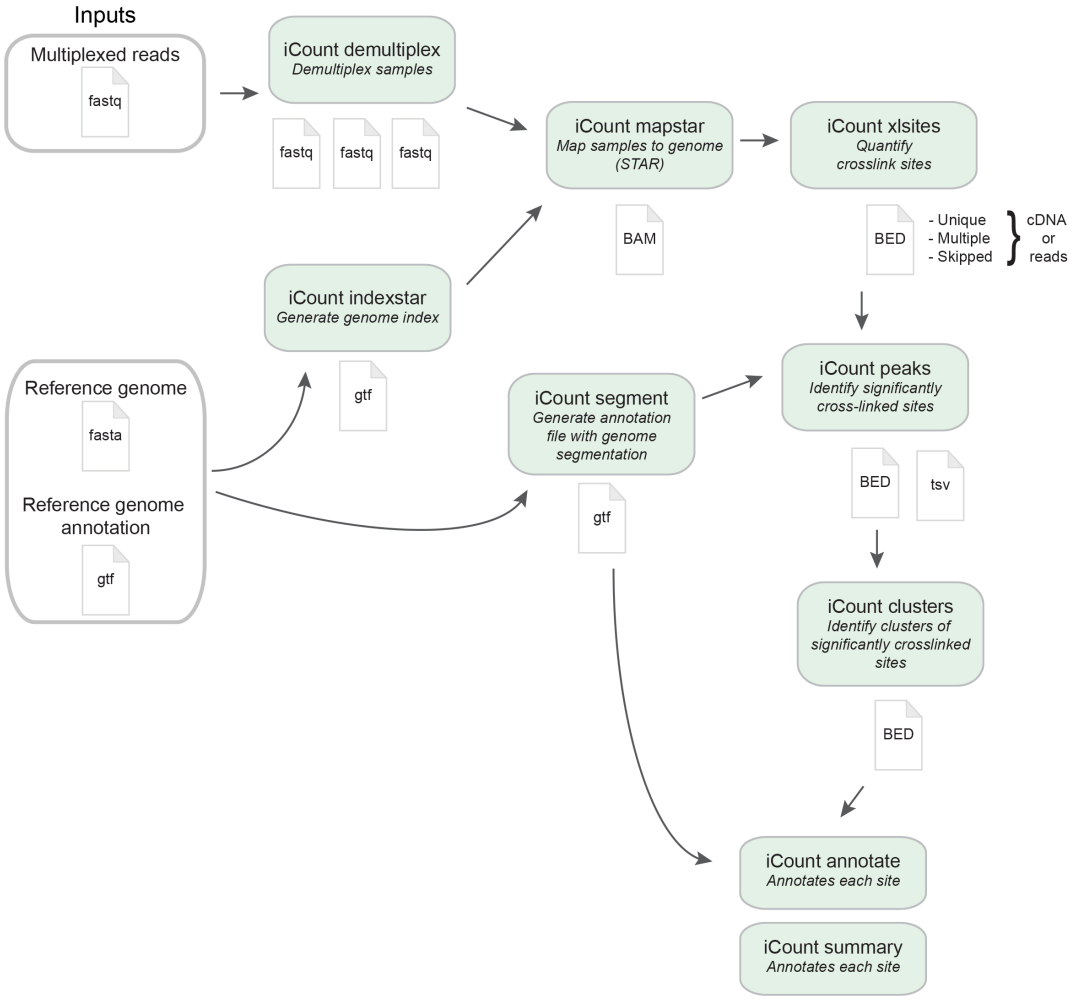


Figure 5.13: iCLIP iCount analysis pipeline. Schematic of the iCLIP analysis using the iCount pipeline (Curk *et al.*, in preparation). Samples are demultiplexed (iCount demultiplex), before mapping to the relevant assembled genome index (iCount indexstar and mapstar). Crosslinked sites are then identified, and sorted based on whether they are unique/mutlimapped or skipped (iCount xlsites). This can be performed based on cDNA molecules or reads, thus enabling estimation of the level of over-amplification that has occurred. Using an annotation file with genome segmentation information (iCount segment), significantly crosslinked sites can then be identified (iCount peaks), and subsequently clustered (iCount clusters). Clusters of significantly crosslinked sites can then be annotated (iCount annotate) and summarised (iCount summary), before downstream biological analysis and interpretation.

in which signal detected in the eGFP negative control at each genomic position is subtracted from the eGFP-p53 signal. It is known that eGFP can co-precipitate with highly abundant transcripts. This analysis allows us to differentiate between true signal and noise derived from the eGFP control. At the time of writing, I am in the process of analysing my iCLIP data using this adapted pipeline.

5.4 The role of RNA in mediating p53 complexes

Traditionally, RBPs and their interactions with RNAs have been considered from a mainly protein-centric perspective. With this viewpoint, proteins bind RNAs and exert a range of functions upon the RNA molecules, thus impacting their stability, localisation, translation and decay. New discoveries have revealed that RNA can be more than a passive passenger in ribonucleoprotein complexes; indeed, RNAs themselves can bind proteins, impacting their function, structure, localisation, and facilitating the formation of multi-protein complexes [11]. However, most protein-protein interaction analysis obtained by pull down and immunoprecipitation strategies lack nuclease treatments. This raises the important question of whether observed complexes represent direct protein-protein interactions, or involve crucial RNA intermediates.

To investigate the role of RNA in facilitating p53 protein complex formation, GFP-p53 was immunoprecipitated from etoposide-treated cells using the eGFP nanobody under mild conditions (150 mM NaCl) to preserve protein-protein interactions. During the second wash step of the IP, half the samples were treated with RNases (+RNase), with this step omitted for the other half (-RNase). IP washes were then completed, proteins eluted from the GFP-trap beads, and eluates subjected to label

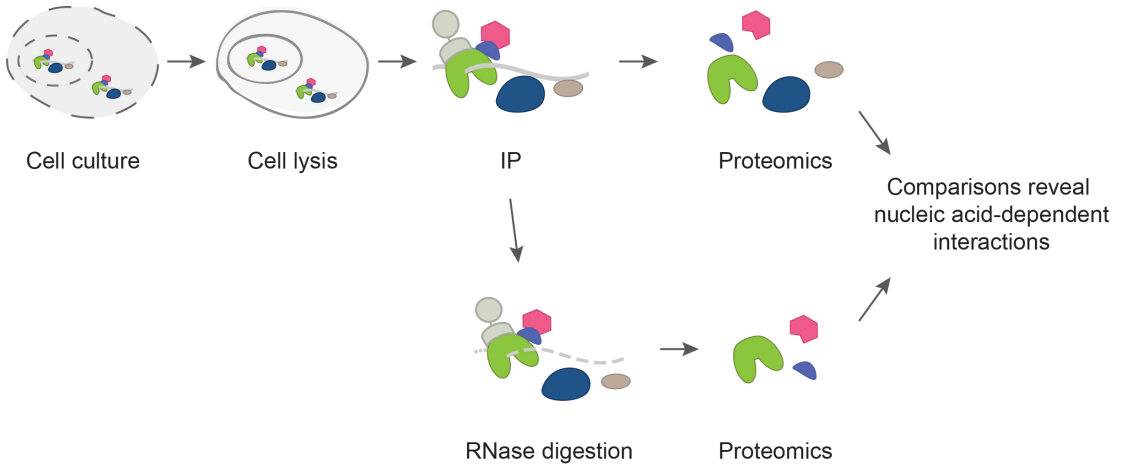


Figure 5.14: RNase dependent protein-protein interaction protocol. To identify RNA-dependent protein interactions, IPs were performed under near physiological conditions. During washing, IPs were either treated with RNases, or not. Resulting IP eluates were subject to label free proteomic analysis. The resulting comparisons then seek to identify proteins whose interactions with the target protein are altered by RNase treatment.

free proteomic analysis (in collaboration with Wael Kamel and Shabaz Mohammed). RNase treatment during the wash steps will remove any protein interacting with p53 via an RNA bridge. Comparisons of +/- RNase proteomes can then be used to evaluate the role of RNAs in mediating protein complex formation (Figure 5.14); distinguishing between direct and indirect protein interactions.

The first attempt at evaluating the impact of RNase treatment on the p53 protein interaction network yielded counterintuitive results. Overall, a higher number of proteins were observed upon RNase treatment (Figure 5.15 A), opposing the hypothesis that digestion of RNA would remove RNA-mediated interactors from the sample. Further investigation revealed that nucleic acids can also be captured by the carboxylate-coated beads used during proteomic clean up, thus disrupting protein binding to the beads [89], and therefore reducing the protein content of the final sample processed via MS. Accordingly, the experiment was repeated, includ-

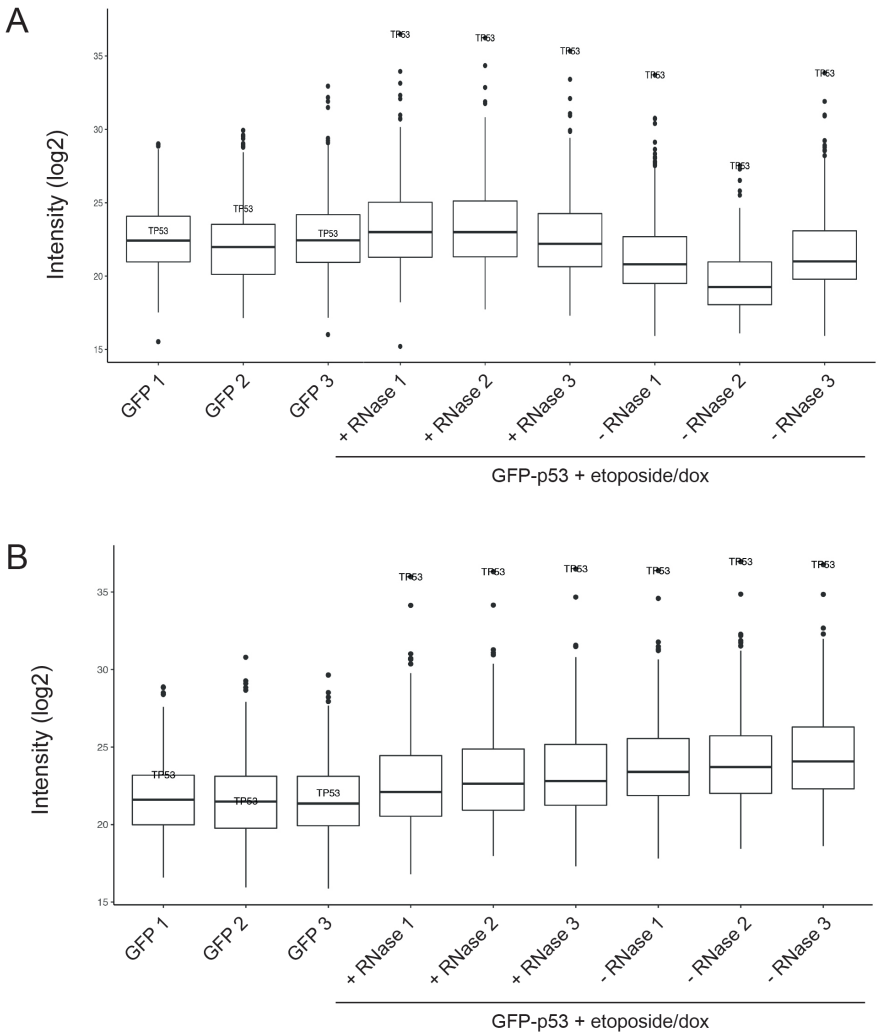


Figure 5.15: Nucleic acids impact MS processing. Boxplots showing the distribution of protein intensities for GFP and GFP-p53 IPs subject to label free mass spectrometry. (A) Initial experiment, where samples not treated with RNase (-RNase), displayed a reduction in the number of proteins observed compared to those treated with RNase (+RNase). (B) Subsequent experiment, where all samples were treated with benzonase after IP and elution, to remove nucleic acids that may disrupt protein processing steps.

ing a benzonase treatment at the elution step to remove nucleic acids that interfere with MS sample processing. This remedies issues with nucleic acids binding the carboxylate-coated beads, and results in an increase in the number of proteins identified without RNase treatment (Figure 5.15 B), and a comparatively reduced number of proteins identified upon RNase treatment.

5.5 Discovering RNase sensitive p53-protein interactions

Protein intensity ratios were calculated between +RNase and –RNase samples, and the significance of these ratios estimated using moderated t-test. LC-MS and initial analysis were performed by Wael Kamel. In order to account for proteins that may copurify with the eGFP tag, proteins identified previously as significantly enriched in the GFP control immunoprecipitates compared to etoposide/doxycycline treated GFP-p53 immunoprecipitates (Chapter 4) were highlighted as 'GFP binders'. Although included in the following volcano plots, GFP binders were removed for all subsequent analyses, and any protein numbers quoted do not include GFP binders (unless stated otherwise).

The resulting volcano plot (Figure 5.16 A) shows fold change (log2) against adjusted p-values (-log10). Coloured points denote proteins that show significant differences between + and – RNase conditions, with yellow and light blue dots representing an adjusted P value < 0.05, and red points denoting a P value < 0.01. GFP binders are highlighted in dark grey. 204 proteins display significant difference between RNase conditions (adjusted P value < 0.05), of which 66 can be termed high confidence (adjusted P value < 0.01), although the volcano plot shows a tendency for most of the non-significant proteins to be enriched in the - RNase sample.

These 204 RNase sensitive proteins were then subject to network analysis (using STRING [113]). The resulting network is shown (Figure 5.16 B), and functional complexes annotated. Network analysis of the proteins that exhibit RNase sensitivity reveals clusters of functionally related protein complexes. Interestingly, many

of these proteins are involved in the initiation and early stages of translation, along with pre-mRNA splicing. This includes 12 of the 13 eIF3 complex members, along with eIF2α, 15 components of the 40S ribosome, and several components of the spliceosome; including 8 small nuclear ribonucloprotein subunits, and 3 out of 7 splicing factor B (SF3B) complex subunits.

This link to RNA biology is emphasised when RBPs are highlighted in the volcano plot. 255 known RBPs [11] are identified across the 2 IP conditions (831 proteins identified in total), which reinforces the high interconnectivity of the p53 pathway with RNA metabolism. 123 of these RBPs are significantly enriched in the –RNase condition, whilst only 1 (MKI67) is significantly enriched in the + RNase condition (Figure 5.17 A). These RBPs are highlighted in the network analysis (Figure 5.17 B). Brannan *et al.* [116] suggest that utilisation of RNase-coupled proteomic strategies, similar to the one employed here, enables the identification of RNA-dependent and independent protein interactions. Brannan *et al.* demonstrated that proteins establishing interaction with RBPs are likely to be RBP themselves, and that this probability increases as the number of RBPs co-immunoprecipitated increases. Hence, the large proportion of RBPs co-immunoprecipitated with p53 bolsters our earlier data showing that p53 is a bona fide RBP.

Enrichment analysis performed using the RNase sensitive protein dataset revealed that proteins annotated with the 'RNA binding' gene ontology term are predominant in this subset of proteins, along with 'nucleic acid binding' and 'mRNA binding'. 'Translation initiation factor activity' and 'translation regulator activity' also appear in the top 10 terms (Figure 5.18 B). Translation and mRNA-associated terms appear in the enriched GO biological process terms (Figure 5.18 A). Enrich-

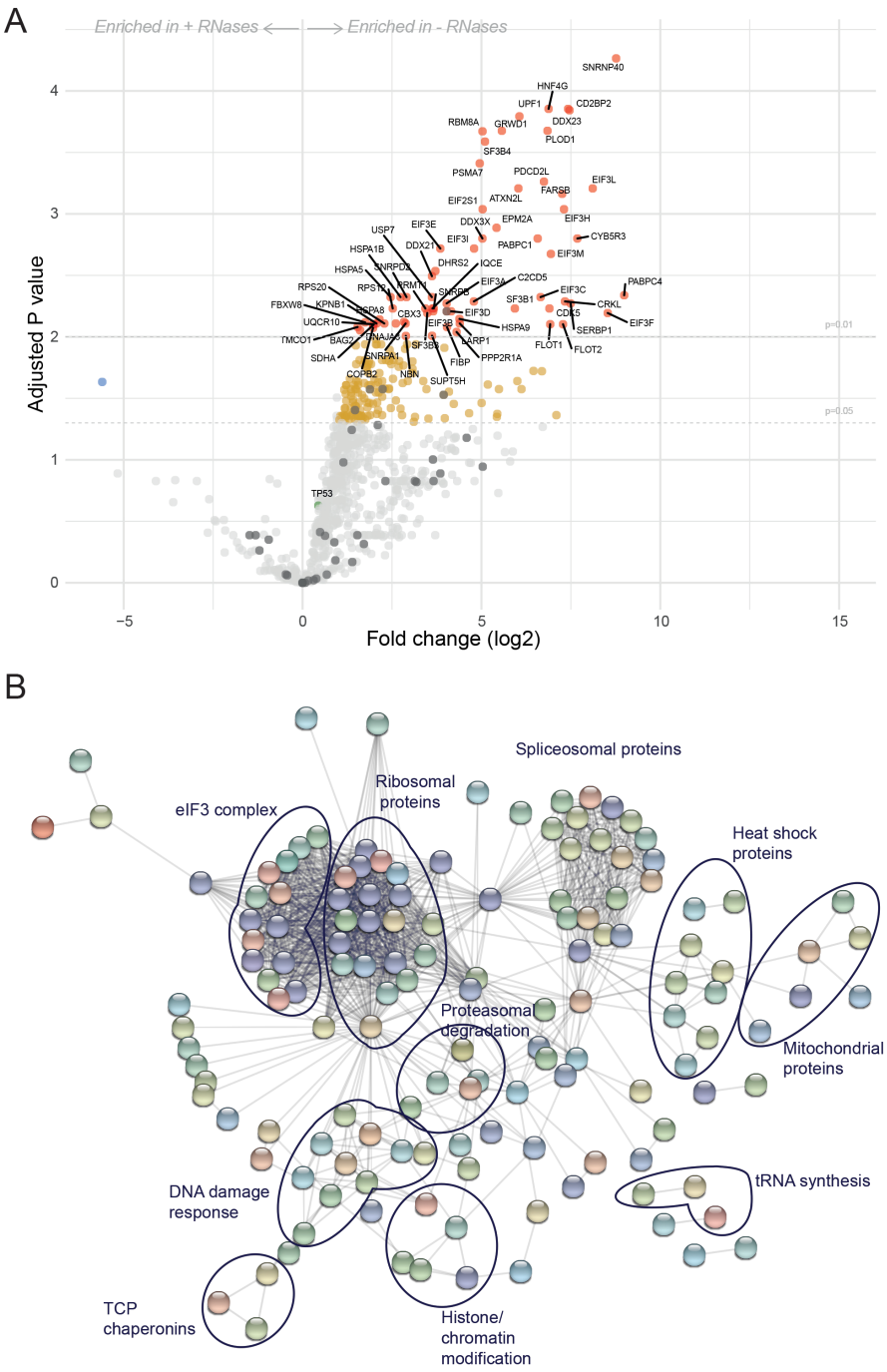


Figure 5.16: Proteins significantly impacted by RNase treatment. (A) Volcano plot plotting log2 fold change against -log10 adjusted P value, for etoposide/dox treated GFP-p53 without RNase treatment, compared to etoposide/dox treated GFP-p53 with RNase treatment. Proteins with an adjusted P value < 0.05 are highlighted in yellow or light blue, and those with an adjusted P value < 0.01 are highlighted in red or dark blue. GFP binders are shown in dark grey, and excluded from subsequent analyses. (B) STRING analysis of proteins significantly impacted by RNase treatment (adjusted P value < 0.05), excluding GFP binders. Functional complexes are annotated.

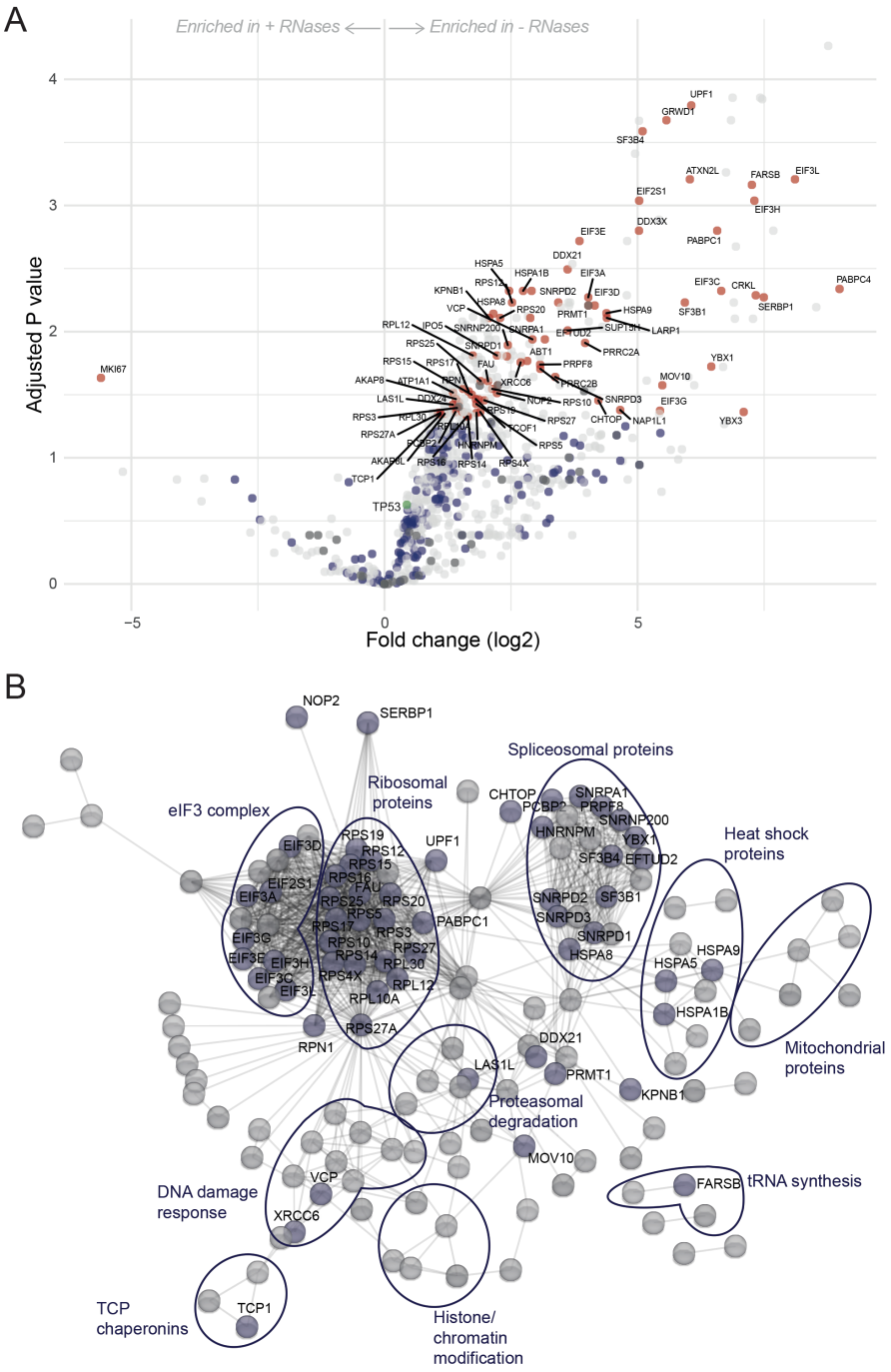


Figure 5.17: RNA binding proteins are impacted by RNase treatment. (A) Volcano plot plotting log2 fold change against -log10 adjusted P value, for etoposide/dox treated GFP-p53 without RNase treatment, compared to etoposide/dox treated GFP-p53 with RNase treatment. Known RBPs are highlighted, with those with an adjusted P value < 0.05 highlighted in red, and those with an adjusted P value > 0.05 highlighted in blue. GFP binders are shown in dark grey, and excluded from subsequent analyses. (B) STRING analysis of proteins significantly impacted by RNase treatment (adjusted P value < 0.05), excluding GFP binders. Known RNA binding proteins are highlighted in blue. Functional complexes are annotated.

ment for a number of early translation-associated events in the reactome pathways analysis supports the idea that these proteins represent components of the preinitiation complex and translation initiation machinery (Figure 5.18 D). Additionally, 'Ribosome', 'Spliceosome' and 'RNA transport' comprise the top 3 significantly enriched KEGG pathways within this subset of proteins. DNA-damage associated pathways can be seen in the KEGG analysis, including 'Non-homologous end joining', 'cellular senescence' and 'cell cycle' (Figure 5.18 C), which represent key p53-mediated responses to DNA damage.

These results suggest that, at least in a proportion of cases, these proteins, interact with p53 in an indirect manner, via an RNA intermediate. Additionally, the RNase sensitivity of multiple mRNA-associated, and early translation associated proteins, suggests these p53-protein interactions may be modulated via an mRNA intermediate. PolyA-binding protein cytoplasmic 1 (PABPC1) and PolyA-binding protein cytoplasmic 4 (PABPC4) are RNase-sensitive candidates, known to bind polyadenylated RNAs, shuttling between the nucleus and cytoplasm, with roles in pre-mRNA processing and the regulation of translation initiation. The presence of PABPC1 and PABPC4 in an RNase-sensitive manner provides additional evidence – complimentary to the initial RIC result – that p53 binds to polyadenylated RNAs, which is likely to be occurring within the cytoplasm. The presence of initiation factors and ribosomes in these RNase-sensitive p53 interactomes strongly suggest that p53 is associated with mRNAs engaged in translation. Taken together, these data evoke an additional layer to p53's cytoplasmic functions.

Importantly, 50 of the 204 RNase-sensitive protein interactors have been previously identified to interact directly with p53 (STRING – experimental evidence or

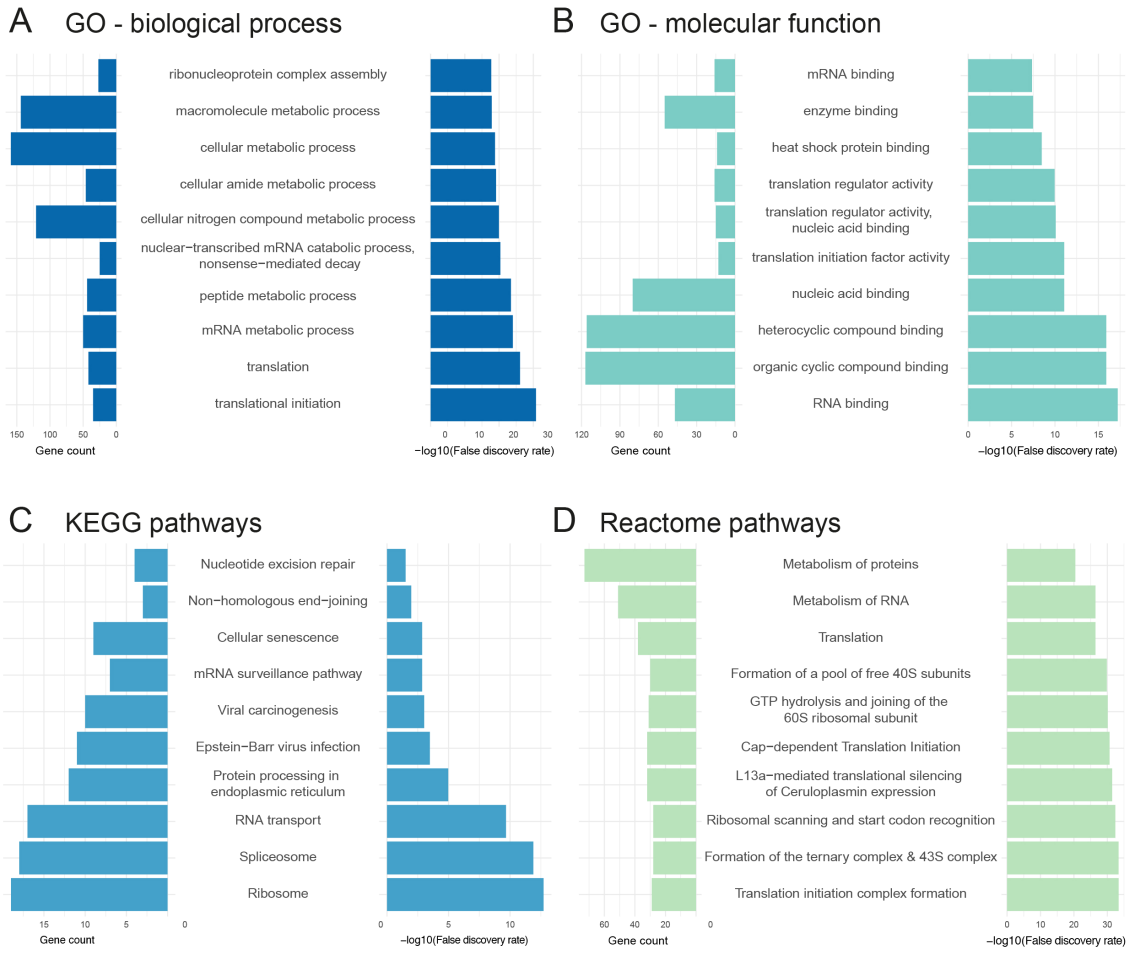


Figure 5.18: RNase sensitive proteins - enrichment analysis. STRING network analysis was performed on RNase sensitive proteins (excluding GFP binders), followed by enrichment analysis. The top 10 terms are shown for the Gene Ontology biological process terms (A) and molecular function terms (B), along with KEGG pathways (C) and Reactome pathways (D). Enrichment analysis was performed against the whole genome, and results are ordered by false discovery rate.

presence in curated database, any confidence level). These proteins are highlighted in Figure 5.19, along with functional protein complexes. The RNase sensitivity of 50 'known' p53 interactors suggests these interactions may be mediated by an RNA intermediate, or concurrent binding to a shared RNA target, in at least a proportion of cases. This highlighted the requirement for inclusion of nuclease treatments in proteomic approaches whose aim is to identify direct protein interactions, and raises questions as to the validity of 'direct' p53-protein interaction data to date. Moreover, the fact that a large proportion of p53 interactions are RNase sensitive further validates previous results, that p53 is an RBP, from an orthogonal angle [116].

5.6 RNase-coupled proteomics reveals uncharacterised RBPs

Examination of RNase-coupled proteomic datasets may provide us with insights into proteins with previously unanticipated or poorly understood roles in RNA biology. An interesting cluster of proteins that is highlighted by this analysis is a group of molecular chaperones; specifically several members of the HSP70 family (HSPA1B, 5, 6, 8 and 9), and their associated HSP40 cofactors (DNAJA1, 2 and 3). HSP70 family members have been documented to associate with translating ribosomes, and bind co-translationally with nascent polypeptide chains in order to facilitate correct peptide maturation, folding and assembly [130]. Additionally, both HSP70 and HSP40 family members have been identified as possessing potential RNA binding activity in RIC and RBDmap experiments [11, 15]. Indeed RBDmap has identified high confidence RNA-binding sites within the heat shock domains of both HSP70 and

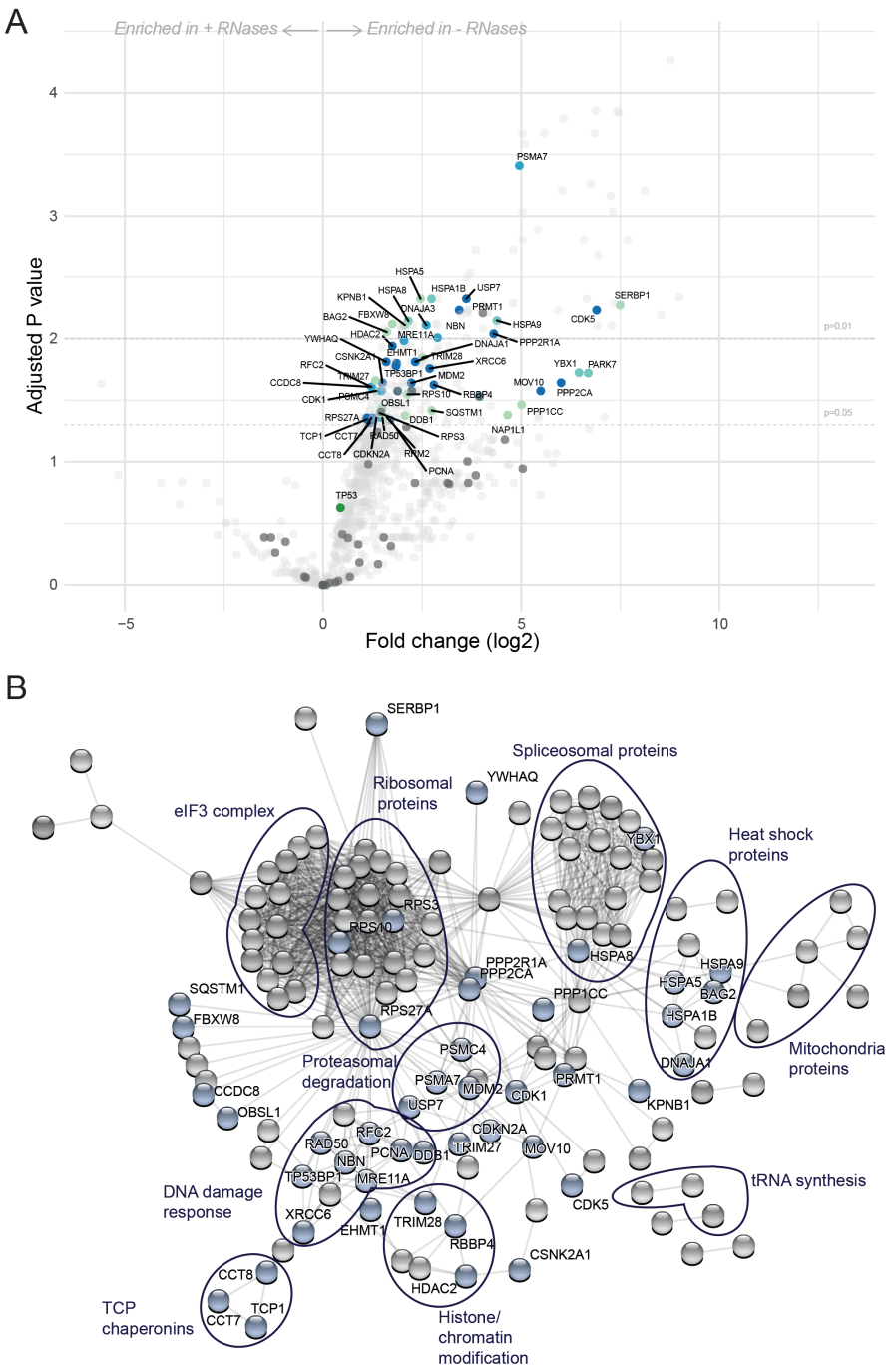


Figure 5.19: Known p53 interactors are impacted by RNase treatment. Volcano plot plotting log2 fold change against -log10 adjusted P value, for etoposide/dox treated GFP-p53 without RNase treatment, compared to etoposide/dox treated GFP-p53 with RNase treatment. Known p53 interactors are highlighted based on their confidence level (from STRING). GFP binders are shown in dark grey, and excluded from subsequent analyses. (B) STRING analysis of proteins significantly impacted by RNase treatment (adjusted P value < 0.05), excluding GFP binders. Known p53 interactors (as defined by STRING) are highlighted in blue. Functional complexes are annotated.

HSP40 proteins (along with HSP90, HSP27 and HSP60 proteins). Moreover, HSP70 has been noted to bind to AU-rich elements in mRNAs, and regulate their stability, independent of its chaperone activities. Identification of heat shock proteins as RNase-sensitive p53 interacting proteins provides additional evidence for their RNA binding activities within cells. Such analysis can, therefore, provide interesting insights into proteins with previously unidentified RNA binding activities.

5.7 RNA independent p53 protein interactions reveal DNA damage induced protein complexes

Consideration of 'non-significant' hits within this dataset, representing proteins that do not exhibit RNase sensitivity, reveal high confidence, directly interacting protein complexes. By adjusting the significance threshold (adjusted P value > 0.95) to highlight proteins that do not exhibit RNase-dependent coimmunoprecipitation with p53, we are able to identify complexes in direct interaction with p53. Using this threshold yielded 38 proteins, which were then subject to STRING network analysis. Figure 5.20 shows the protein interaction network for the RNase insensitive proteins, with p53 included, and functional complexes annotated.

Network analysis (Figure 5.20) validates 6 previously identified direct p53 interactors, including 3 low confidence interactions. This network demonstrates a high degree of connectivity, with 19 out of the 38 nodes connected in a main central cluster (Figure 5.20). This central group of proteins contains members whose functions provide key insights into cellular responses to DNA damage. The cluster of histone proteins play key roles in the remodelling of chromatin after DNA damage [131],

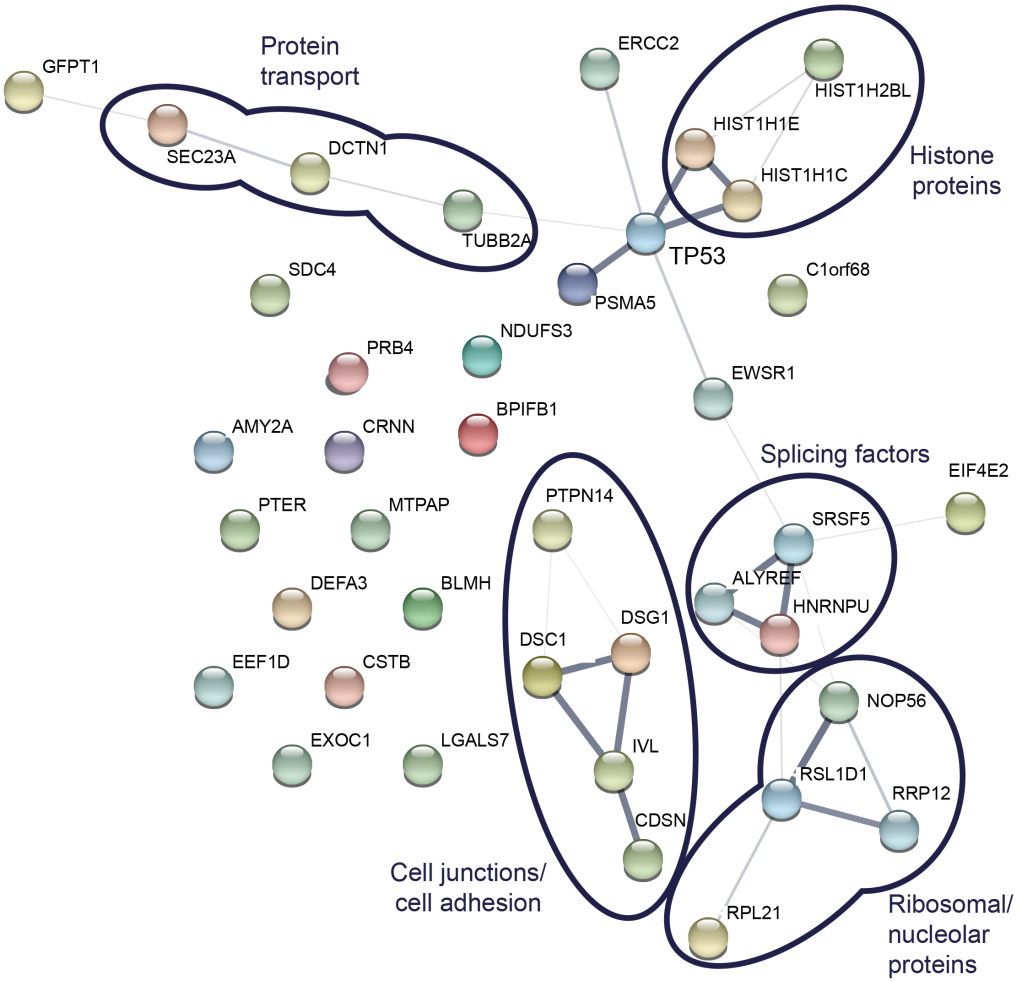


Figure 5.20: RNA independent p53 interactions. STRING analysis of proteins exhibiting RNase insensitive/RNA independent interactions with p53 (adjusted P value > 0.95). Functional complexes are annotated, and interaction confidence is denoted by line width.

with HIST1H1E and HIST1H1C marking sites of DNA damage, through ubiquitin signalling [132]. The splicing factor cluster, along with EWSR1, are proposed to faciliate the DNA damage response by eliminating R-loops formed at DNA damage sites [133]. The presence of a ribosomal/nucleolar protein cluster is likely a result of DNA damage-induced nucleolar stress, which in turn induces the p53 pathway [134]. The protein transport cluster facilitates intracellular protein transport, with movement of proteins required to mount a successful DDR [135]. Another interesting member of the central network is EIF4E2, also known as 4EHP, has been shown to facilitate induction of a DNA damage-induced translational arrest, via competitive binding to the 5' mRNA cap [133].

Enrichment analysis was also performed (Figure 5.21). Due to the relatively low number of identified proteins, enrichment results were only available for GO biological process terms (Figure 5.21 A), and reactome pathways (Figure 5.21 B). The enrichment analysis corroborates the presence of multiple DNA damage associated proteins, with several DNA damage, apoptosis, senescence, and cell cycle arrest related reactome pathways present (Figure 5.21 B). The GO biological process terms also reflect some of the core network clusters described above, including protein-DNA complex assembly and chromatin assembly terms (Figure 5.20 A).

A proportion of nodes within this network are connected to neither p53, nor one another (14 out of 38). The role of many of these proteins in the DDR is unclear, and they have no experimental or database-based evidence of interaction with p53 in STRING. Proteins from within this group include 2 mitochondrial proteins (MTPAP and NDUFS3), 2 proteins associated with apoptosis (CRNN and LGALS7, which is incidentally, a p53 target gene), and EEF1D. These proteins may, therefore, repre-

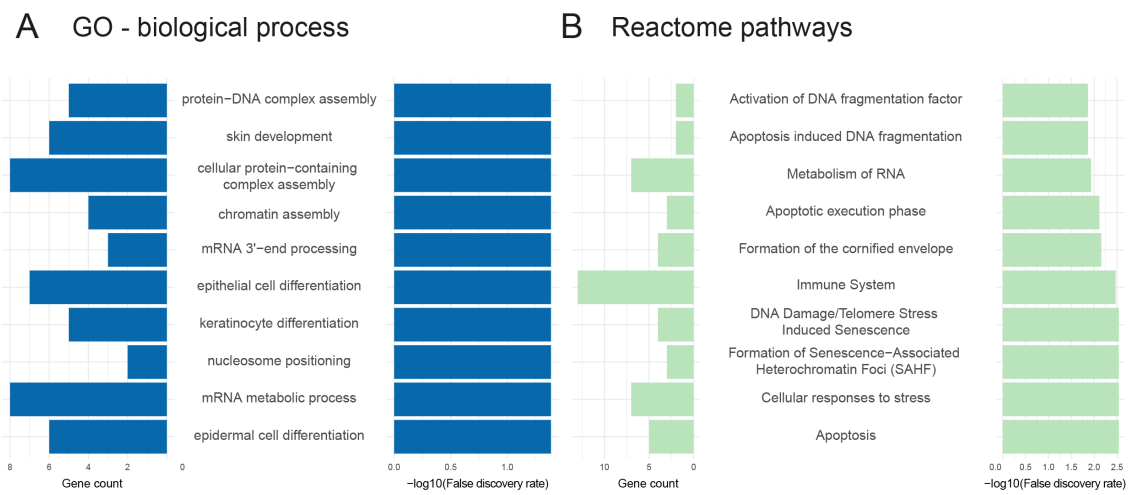


Figure 5.21: RNA independent p53 interactions. STRING network analysis was performed on RNase insensitive proteins, followed by enrichment analysis. The top 10 terms are shown for the Gene Ontology biological process terms (A) and Reactome pathways (B). Enrichment analysis was performed against the whole genome, and results are ordered by false discovery rate.

sent novel or poorly characterised DDR effectors that require further investigation.

To summarise, proteomic analysis of GFP-p53 immunoprecipitations either with or without RNase treatment, identified over 800 proteins. Of these, 204 exhibit significant sensitivity to RNase treatment. This highlights the importance of RNAs in the formation of protein complexes [136], with even well characterised direct p53 interactors, such as MDM2, exhibiting a degree of RNase sensitivity. This also corroborates our results indicating p53 binds to RNA in living cells. Indeed, the RNase sensitivity of translation initiation factors, 40S ribosomal components and PABPC1 further corroborates previous *in vitro* works, and our RIC data, suggesting p53 binds to protein coding mRNA transcripts. Interaction with a large number of RBPs further emphasises the link between p53 and RNA metabolism, whilst the RNA dependence of a large number of these interactions provides orthogonal evidence for p53's function as an RNA binding protein [116]. Probing of the RNA independent p53 interactions reveals an interconnected network of p53 interacting

proteins, along with important players in the DNA damage response. Poorly characterised DNA damage repair/response proteins are identified, that require more intensive study. The relatively large number of RNase sensitive proteins, compared to a much smaller number of RNase insensitive proteins suggests RNA plays a key role in the DDR, facilitating the formation of protein complexes under these cellular stress conditions.

5.8 The RNA binding domain of p53

In the quest to characterise p53's RNA binding activity, previous *in vitro* works converged on the consensus that the C-terminal domain of p53 is likely responsible for its observed RNA binding activities. This conclusion was obtained through the use of either C-terminal truncated recombinant p53, C-terminal p53 fragments, or antibodies targeting the C-terminus of p53, in predominantly acellular systems. The C-terminus has been shown to bind ssRNA, dsRNA, ssDNA and dsDNA, all in a seemingly non-specific manner [72, 75, 76]. However, our data revealed that p53 interacts with a defined subset of RNAs, suggesting that its interactions with RNA exhibit some degree of specificity.

In silico approaches to predict the RNA binding activity of proteins have previously relied on the presence of recognised, classical RNA binding domains. However, it has become increasingly apparent that RBPs often do not posses such domains, and are still able to interact efficiently with RNA [15, 13, 8, 14, 41] As such, predicting the RNA-binding propensities of non-classical RBPs remains challenging.

Our collaborators (Bernd Fischer and Ni Shuai) developed RBDetect, which utilises RBDmap experimental data [15] to predict the probability of RNA binding

within protein sequences. RBDmap utilises successive rounds of oligo(dT) capture, interspersed with proteolytic digestion of RNA bound proteins, to enrich for protein peptides engaged in RNA binding in living cells. Peptides are then subject to tryptic digestion and quantitative proteomics. The nucleotide(s) crosslinked to the peptide makes identifying bound peptides extremely challenging, due to aberrant and heterogeneous mass shift [15]. However, as the first protease cleaves less frequently than trypsin, the latter would liberate a nucleotide-crosslinked peptide and likely, and an unmodified tryptic peptide. Identification of these unmodified peptides are used as tags to reconstitute, in silico, the original peptide in contact with the RNA. RBDetect utilises a machine learning approach to learn from the sequence properties of RNA-bound and RNA unbound peptides, and predict RBDs in protein with uncharacterised RNA binding domains.

RBDetect was used to predict which regions of the p53 protein might exhibit RNA binding capabilities. Across the p53 protein there are 5 regions with increased probability of binding RNA (Figure 5.22 A, probability > 0.6 for at least 5 consecutive amino acids). Of these regions, 4 map to the core DNA binding domain (amino acids 103 – 107, 155 -161, 186-187, 267-276). An additional region of increased probability is observed within the oligomerisation domain (amino acids 323-330), and a final cluster of 3 regions within the C-terminal domain (spanning amino acids 355-371). When plotted onto a three-dimensional structure of p53 combining available structures and homology modelling for the regions that have not been crystallised, the three C-terminal sites come together into close three-dimensional proximity; forming what could represent a binding pocket (Figure 5.22 B). This supports the notion that the C-terminal domain of p53 may mediate RNA binding activity, using

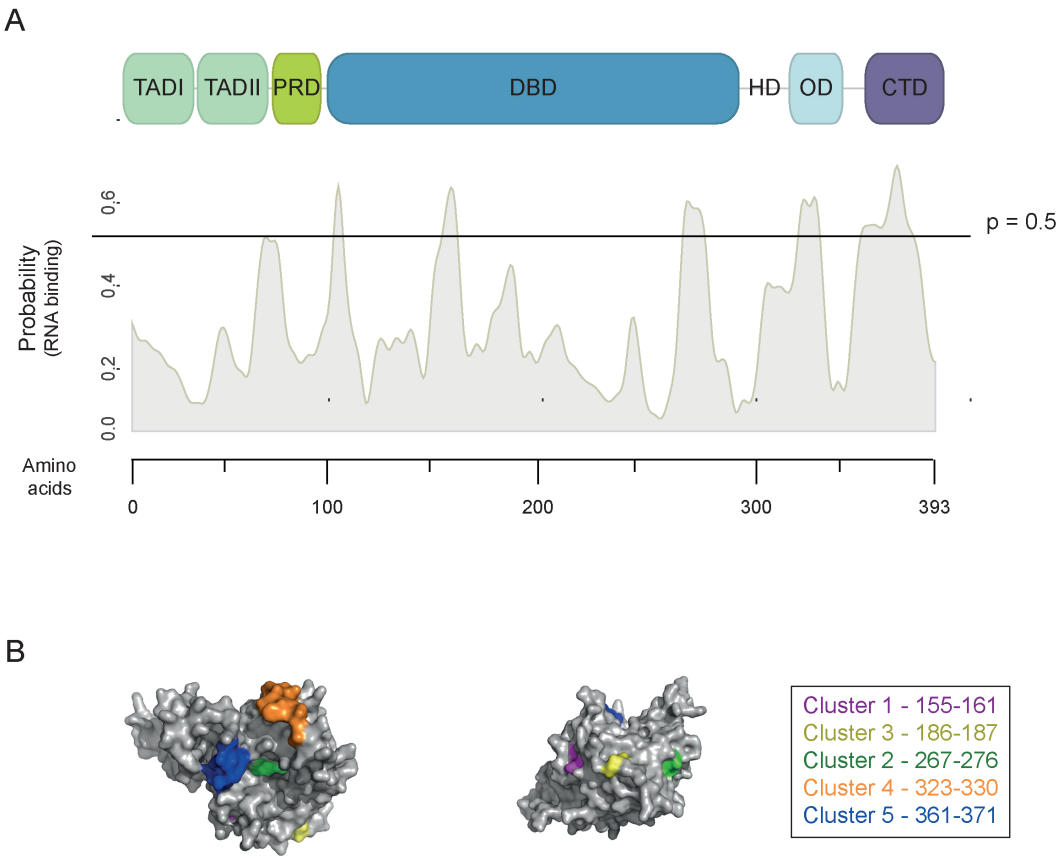


Figure 5.22: Prediction of p53's RNA binding sites. (A) Schematic representation of the p53 protein domain architecture, with the predicted probability of RNA binding capabilities across the protein. Predictions using machine learning based on RBDmap data. Predicted regions of RNA binding capability are highlighted on the p53 protein structure (B) using homology modelling.

a predictive model that takes into account the propensity of non-classical protein sequences to mediate protein-RNA interactions. It should be noted that due to the presence of intrinsically disordered regions (IDRs), structures of the CTD are still lacking, and this model considers these regions in the homology modelling. Hence, these results, although informative, should be taken with caution.

5.9 Discussion

5.9.1 p53 interacts with a defined subset of cellular mRNAs

The p53 tumour suppressor has previously been shown to exhibit RNA binding activity in acellular and *in vitro* systems. Whilst demonstrating the biochemical propensity of p53 to bind RNAs, these studies provide little evidence regarding physiological interactions between p53 and RNA, and their biological significance. Prior attempts to identify p53's RNA interaction partners in living cells proved problematic, due to the coimmunoprecipitation of contaminating RNPs when using conventional IgG antibodies [104], as evident in my initial work in the previous chapter. This strongly affects our ability to study p53 under physiological conditions, due to the presence of unwanted protein and RNA species that will contaminate our immunoprecipitations, and may impact experimental read out. As such, the Hek293 GFP-p53 inducible cell system was used to enable further characterisation of p53-RNA interactions in living cells; due to the capacity of the GFP-trap system to produce extremely clean immunoprecipitations, as shown in Chapter 4. We have exploited this approach to study p53 protein-protein and protein-RNA interactions using CLIP combined with next generation sequencing, and proteomic analysis.

CLIP-seq was used to isolate and identify RNAs interacting with p53, in either etoposide or nutlin treated cells. CLIP-seq revealed 38 candidate transcripts bound to p53 in both nutlin and etoposide treated cells, of which 33 were classified as 'high-confidence'. All 38 transcripts were mRNAs. These results first confirm experiments in Chapter 4, demonstrating, by multiple orthogonal approaches, that p53 is endowed with RNA-binding activity in living cells. RIC, utilised in Chapter

4, employs oligo(dT) purification of polyadenylated RNAs, which enriches predominantly for mRNAs (approximately 70% of the RNA eluted in HeLa cells [10]), thus supporting p53's ability to interact with mRNAs. Secondly, this contradicts the lack of RNA-binding specificity proposed by several *in vitro* studies using purified p53. Indeed, p53 recognises a rather limited subset of RNAs. This is in accordance with other *in vitro* works showing that p53 interact with a small selection of mRNAs [80, 78, 79, 77]. Additionally, RNase-coupled proteomics reveals a large number of ribosomal components and translation initiation factors, along with the cytoplasmic poly(A) binding protein PABPC1, as RNA dependent interaction partners of p53. These are all key components of the protein synthesis apparatus, and their RNA-dependence supports the notion that p53 is binding mRNAs that are likely engaged in protein synthesis.

p53 has been linked to the direct regulation of specific RNAs *in vitro*. Indeed, the CLIP-seq dataset enabled validation of several *in vitro* works describing p53 protein interacting with the RNAs of key regulators MDM2 and MDM4, along with p53 itself. Although the function of these interactions have not yet been elucidated, *in vitro* works focussing on p53 binding to its own mRNA demonstrated translational repression of the *p53* mRNA upon p53 binding. Mosner *et al.* suggested that p53's RNA-RNA annealing activity may stabilise secondary structures within the *p53* mRNA 5' UTR, or enable refolding of the mRNA into a more thermodynamically stable form [80]. This in turn would lead to reduced translation of p53 mRNA. Other *in vitro* works, although not elucidating a precise mechanism, support binding of p53 to the 5'-UTR of *MDM4* [79], *CDK4* [78] and *FGF2* [77, 137] mRNAs. In contrast, p53 is proposed to repress *MDM2* and *p21/Cdkn1a* transla-

tion via an alternative mechanism, not involving the 5'UTR, but instead binding to target transcripts' coding sequences [81]. Zaccara *et al.* [138] showed significant uncoupling of transcription and translation in MCF7 vector cells treated with nutlin or doxorubicin. 70% of differentially expressed genes (nutlin vs doxorubicin) exhibited uncoupling of transcription and translation [138]. Taken together, all these data support the hypothesis that p53 binding to mRNAs regulates their translation, adding an additional layer of complexity to p53's regulation of gene expression. Analysis of my iCLIP data will reveal if p53 recognises any sequence motifs or secondary structure, and if it does indeed interact with the 5'-UTR of target mRNAs in living cells.

Of the 38 mRNAs bound by p53 in nutilin- and etoposide-treated cells, 17 are linked to p53 in the literature. However, only 3 of these shared candidates (*p53, MICALL1 and MDM2*) are core p53 transcriptional targets (as defined by Fischer *et al.* [123]).

Interestingly, mRNAs encoding histone modifiers are also highlighted as high confidence targets of p53. Lysine Methyltransferase 2D (*KMT2D*) and lysine demethylase 5C (*KDM5C*) mRNAs are both identified in our p53 CLIP-seq dataset. Interestingly, both histone modifiers have antagonistic roles as methylation and demethylation of lysine 4 of histone 3 drive transcriptional activation or repression, respectively. p53 gain of function (GOF) mutants (R248Q, R249S and R273H specifically) upregulate KMT2D along with other chromatin regulatory genes, to alter histone methylation and acetylation patterns genome-wide and, as a consequence, dysregulate gene expression [85]. Transformation/transcription domain associated protein (TRRAP) is a member of the PIKK family of proteins that has been shown to be

a component in a range of histone acetyltransferase complexes. TRRAP plays a role in transcriptional responses, including p53 and MYC responses, through its recruitment of histone acetyltransferase complexes to chromatin. Strikingly, *TRRAP* mRNA is also present in eGFP-p53 IPs. Taken together, these data strongly suggest that p53 may regulate histone modifiers, which could contribute to p53's function as a transcription factor.

Another group of high confidence p53 targets are the mRNAs encoding nucleoporin 98 and 153 (*NUP98 and NUP153*), which are both key regulators of RNA and protein trafficking across the nuclear pore complex (NPC). Exportin 5 (*XPO5*) is another mediator of the nuclear/cytoplasmic transport of pre-miRNAs and tRNAs. XPO5, NUP98 and NUP153 are therefore all involved in nucleocytoplasmic transport of macromolecules. These functions can therefore impact upon gene expression in multiple ways. Transcription factors, such as p53, and NF-KB, require transport into the nucleus upon cellular stress in order to execute their transcription factor functions, which is reliant upon the nuclear pore complex. The NPC, and XPO5, are also involved in the transport of RNA species from the nucleus to cytoplasm. Interestingly, upon DNA damage, NUP153 is phosphorylated by the ATM-activated AKT kinase, which enhances its interaction with XPO5, facilitating increased nuclear export of pre-miRNAs [139], which may facilitate miRNA-mediated responses to DNA damage [140, 141]. Along with their role in mediating macromolecule transport, nuclear pore complexes have been implicated in the regulation of gene expression, with both Nup98 and Nup153 highlighted as influencing both gene activation and repression [142], including several p53 response genes [143], through largely unclear mechanisms. Binding of p53 to *NUP98, NUP153* and *XPO5* could therefore

represent mechanisms by which p53 is able to alter the availability of components involved in nuclear-cytoplasmic trafficking of macromolecules, including miRNAs or transcription factors, which are both crucial for p53 response to cellular stress.

In summary, CLIP-seq reveals mRNA networks that provide exciting insights into the potential role of p53 in the post-transcriptional regulation of p53-driven gene expression. The striking enrichment of p53 mRNA and mRNAs encoding several p53 regulators in eGFP-p53 IPs suggests the possibility of extensive post-transcriptional autoregulatory feedback loops that may enable fine-tuning of the p53 response, while enabling rapid responses to cellular stress. Additionally, p53's binding to the transcripts encoding histone modifiers and nucleoporins represents a potential post-translational regulatory layer with ability to control p53-driven transcriptional activation. Further work is therefore required to unpin the mechanisms and biological function of these interactions.

In order to experimentally assess the role of p53 in translational regulation, I will apply ribosome profiling to my GFP-p53 system, in the presence and absence of p53, along with comparison of genotoxic versus nongenotoxic activation conditions. Ribosome profiling reveals RNA regions protected from mild RNase digestion by the ribosomes, termed ribosome-protected fragments (RPFs), yielding ribosome footprints in a transcriptome-wide manner [144]. Ribosome profiling thus reveals which RNAs are actively translated in living cells, and can be used to assess the role of p53 in translational regulation. Combination with RNA-seq will enable me to elucidate if the CLIP-seq p53 target RNAs are translationally regulated in the presence or absence of p53 [106].

5.9.2 Disparities in RNA-binding activity in genotoxic versus nongenotoxic p53 activation

Cells were treated with two different p53 inducing stimuli – nutlin and etoposide – to enable comparison and overlap of p53 bound transcripts under different conditions, thus identifying a 'core' p53 post-transcriptional network, as has recently been performed with its transcriptional programme [123].

The chemotherapeutic agent nutlin-3a is a cis-imidazoline analogue, which specifically disrupts the p53-Mdm2 interaction, and thus liberates p53 from Mdm2-mediated repression. This disruption results in the non-genotoxic stabilisation of p53, and activation of the p53 pathway [102, 145, 146]. Nutlin treatment therefore represents a highly specific mode of p53 stabilisation, with very few "off target" impacts on gene expression, as it does not trigger parallel stress-related pathways [147].

Etoposide, on the other hand, is a chemotherapeutic topoisomerase II inhibitor, which induces single and double stranded DNA breaks. This leads to stabilisation of p53 via DNA damage associated ATM kinase-mediated pathways, invoking Ser15 phosphorylation, and significant post translational modification of p53 [147, 110]. This genotoxic stabilisation of p53, and activation of the p53 pathway, therefore induces parallel DNA damage associated transcriptional programmes [147, 109]. Indeed, Catzione *et al.* observed the expression of significantly more target genes in response to etoposide when compared to nutlin, in primary human fibroblasts [147]. Of these, a large number were deduced to be due to DNA damage-dependent, p53-independent mechanisms, including induction of the NFK-B pathway in response to DNA damage [147].

Therefore, the genotoxic (etoposide) and non-genotoxic (nutlin) activation of

p53 differ in terms of their impacts both upon p53 itself, through regulatory PTMs, and downstream of p53. I hypothesise that the large discrepancy in the number of p53-bound transcripts, and the number of reads per gene, between etoposide and nutlin treated cells are due to these intrinsic differences in p53 activation. As etoposide appears to induce a wider range of transcriptional responses, as opposed to the highly specific induction of the p53 pathway by nutlin, the increased p53-bound transcripts may be due to an increased availability of target RNAs. This hypothesis will be tested by RNAseq analysis of total RNA purified from nutlin- and etoposide treated cells. Alternatively, the differential RNA-binding activity may be due to the PTM status of p53. Etoposide has been shown to induce not only the crucial Ser15 phosphorylation of p53, but also extensive modifications across the protein [147, 110], whilst nutlin treatment appears to induce minimal PTMs [147]. The PTM status of p53 under various cellular stresses has been extensively studied, although their physiological roles remain unclear. What is established is that a level of post translational modification is required to stabilise p53, and activate its downstream pathways, under a range of cellular stress conditions [56, 148, 149]. Previous *in vitro* works proposed that acetylation [87, 76] of the C terminal domain of p53 abrogates p53-RNA binding in acellular conditions. It is therefore possible that etoposide-induced changes to p53's PTM status may modulate p53 RNA-binding activity, compared to the relatively unmodified nutlin-induced p53. To evaluate whether a difference in PTM status may be responsible for the differences in RNA binding, I will isolate GFP-p53 from nutlin- and etoposide-treated cells and analyse its PTMs by proteomics, in collaboration with Professor Shabaz Mohammed. Utilisation of the range of PTM specific p53 antibodies available would also enable decoding of

the PTM status of GFP-p53 within our cellular systems. Once I have established the differences in p53 PTM status, I will utilise the Hek293 Flp-In TREx system to introduce site specific mutations, that will either imitate or impede residue modification. The RNA binding activity of these GFP-p53 mutants can then be assessed to decode the role of PTMs in p53's RNA binding activity, under these conditions. It is also reasonable to suggest that alterations in cofactor recruitment and availability between the two treatment conditions could explain the significant differences in RNA binding observed.

5.9.3 Further analyses to reveal the non-coding RNA network of p53

Although the majority of unprocessed and GFP-normalised reads from the CLIP-seq dataset correspond to protein coding genes, the removal of multi-mapping and spliced reads during processing eliminated a number of other transcript types, including various noncoding RNAs. In the future, I will revisit the analytical workflow to examine these RNAs, especially given the increasing awareness of the role that ncRNAs can have in the p53 pathway.

Through its function as a transcription factor, p53 can regulate the transcription of a range of noncoding transcripts, including lncRNAs and microRNAs. These p53 responsive transcripts can act as 'effectors' of the p53 response [150, 151, 84]. One such example is the tumour suppressive *MEG3* lncRNA (Maternally Expressed Gene 3), which is downregulated in multiple tumour types [85]. *MEG3* has been shown to directly interact with p53, reportedly through the core DNA binding domain of p53, which increases p53's stability, and subsequent transcriptional activation of

targets such as SESN1 and TGFA [85, 86]. This represents a clear example of RNAs exerting regulatory functions upon their protein interaction partner.

Another interesting example of a p53 responsive noncoding RNA is the *DINO* (Damage Induced Noncoding) lncRNA, which lies just upstream of *CDKN1A*. *DINO* is induced in response to doxorubicin-associated DNA damage, upon which it is shown to bind p53 via its C-terminal domain. This facilitates p53 stabilisation and induction of p53 target genes such as CDKN1A, thus providing a feed forward loop for the amplification of p53 response to DNA damage [83, 152]. Indeed, ncRNAs have been shown to play key roles as regulators in the DNA damage response [141, 140]. Other ncRNAs, such as *LINC-ROR , miR-200a, miR-192, miR-194, miR-143, miR-145* and *miR-34a* [141, 140] have been demonstrated to form feedback loops, themselves being p53 responsive, and subsequently impacting the stability of p53, generally through action on Mdm2/Mdm4. NcRNAs can therefore represent a further mechanism by which the p53 response can be modulated or fine-tuned, and as such, discovery of ncRNAs within our p53 CLIP-seq database would be incredibly interesting. Noncoding RNAs represent a mechanism by which the cell can quickly and dynamically respond to cellular conditions, such as DNA damage, as they are not converted into protein products.

5.9.4 iCLIP data will reveal p53's binding sites within target RNAs

The CLIP-seq approach used differs from iCLIP, as it does not utilise partial fragmentation of protein-bound RNAs. This implies that CLIP-seq is a robust method to identify p53 RNA targets but does not provide information about the exact

cis-acting elements that it recognises within the RNA. It is possible to utilise computational approaches to take advantage of the propensity of the RT to prematurely terminate reverse transcription at the protein crosslink residues, and use this to predict an approximate footprint for binding [153, 154]. A limitation of this approach is that the RT would also terminate at the binding sites of other RBPs sitting on the same RNA, and accompanying p53 throughout the immunoprecipitation.

To circumvent the lack of binding-site resolution in the CLIP-seq experiment, I have utilised iCLIP, which provides single nucleotide resolution information on the binding sites of a given RBP [95]. After extensive optimisation, cDNA libraries were generated and sequenced. During the preparation of this thesis, these libraries were being analysed using the iCount pipeline. The iCLIP dataset will allow us to further validate the cellular targets identified by CLIP-seq, determine which regions of the mRNA are recognised by p53 (e.g. 5' UTR) and learn about the binding preferences of p53, through the use of algorithms to identify particular sequence motifs or structures at the binding sites [117]. While p53 was proposed to bind RNA nonspecifically *in vitro* [87], our CLIP-seq data strongly indicates selectivity in p53's RNA interaction partners. This suggests p53 may recognise primary sequence or secondary structures within target RNAs.

Once the iCLIP data has been analysed, and any binding preferences or specificities have been interpreted, the aim is to further investigate the potential biological functions of the GFP-p53 interaction with RNA. Sequence or structural binding specificity does not necessarily equate to biological function [7]. Indeed, non-specific interactions can have important functional consequences. It is therefore critical to further investigate the potential physiological roles of these interactions (see Discus-

sion).

5.9.5 RNA is a central mediator of p53 interactions

The propensity of p53 to form extensive interactions with other proteins is essential not only for its regulation, but also for its ability to exert its tumour suppressor functions. Under varying *in vitro* and cellular conditions, and using a huge range of methodological approaches, hundreds of p53 interaction partners have been identified, with over 990 interactions denoted in the BioGrid database [155]. However, as a DNA and RNA binding protein, the presence of nucleic acid intermediates may prove problematic in immunoprecipitation-based approaches.

RNase-coupled proteomics was used to evaluate the role of RNA in p53 protein complex formation. Over 200 proteins were identified as being RNase sensitive; showing significantly different protein intensities in IPs with and without RNase treatments. These proteins likely represent proteins forming concurrent interactions with p53-bound RNAs, and provide valuable insight into the RNA species bound by p53. Analysis of these proteins reveals enrichment for eukaryotic initiation factors, the 40S ribosome, the spliceosome and mRNA-related metabolic pathways. This correlates well with the obtained CLIP-seq and RIC results, which suggest p53 binds mainly protein-coding, polyadenylated transcripts; suggesting a potential role in the regulation of mRNAs, as previously postulated in *in vitro* works.

It is important to note that presence in the RNase sensitive dataset suggests that at least a proportion of the p53-protein interactions are facilitated by an RNA intermediate, but this is obviously not the case for all interactions, all of the time. For example, we know MDM2 interacts directly with p53 as its primary regulator,

but our data demonstrates MDM2 exhibits a degree of RNase sensitivity, suggesting a proportion of these interactions involve RNA intermediates. It has previously been shown that MDM2 can bind RNA including the *p53* mRNA [156], and it is therefore plausible that a portion of the MDM2-p53 interactions are mediated by a shared RNA target; potentially the *p53* mRNA itself.

The presence of a large number of RBPs in the p53 protein interaction dataset provides further evidence of a high degree of interconnectivity between the p53 pathway and RNA metabolism. The RNase sensitive nature of a large number of these RBPs further bolsters the notion of p53 being a bona fide RBP [116].

5.9.6 Prediction of p53's RNA binding domain

Exploration of the RNAs bound by p53 accounts for half of the interaction interface. The remaining component requiring consideration is the p53 RNA binding interface; namely which protein domain and amino acid residues are responsible for interactions in living cells. A machine learning approach, trained on RBDmap data, enabled prediction of amino acids with increased RNA binding propensity. Sites within the core DNA binding domain, oligomerisation domain, and C-terminal regulatory domain were identified as having increased probability of RNA binding.

p53 contains two distinct nucleic acid binding domains, within the C-terminal domain and the core DNA binding domain. Previous works highlight the C-terminal domain as the main candidate for p53's RNA binding activities, with observed non-specific ssDNA, nonlinear DNA and RNA binding occurring within this region [52]. Truncation or post translational modification of the CTD leads to disruption of p53's RNA binding activities *in vitro* [87]. However, there still remains conflicting

data as to whether the CTD or core DBD mediates p53's interactions with RNA. Indeed, even recent work addressing p53's interactions with noncoding RNAs yield conflicting results, with the *MEG3* lncRNA and *DINO* lncRNAs binding to the DNA binding domain and C-terminal domain, respectively [85, 83].

In order to experimentally elucidate the domain responsible for RNA binding in our systems, we will generate p53 truncation mutants to evaluate whether RNA binding is abrogated. The Flp-In T-REx system enables relatively easy generation of stable, inducible cell lines, and as such C-terminally truncated p53 expressing lines could be generated. It may be possible to narrow down the exact binding sites through a series of increasing length truncations, or even specific amino acid alterations to highlight more precise binding interfaces.

To summarise, I have been able to establish, optimise and utilise CLIP-seq and iCLIP techniques to generate the first transcriptome-wide dataset evaluating p53-RNA interactions in living cells, utilising the GFP-p53 system previously developed. Whilst iCLIP data is still under analysis, CLIP-seq reveals a set of core p53 RNA targets, which are present under both genotoxic and non-genotoxic stress conditions. These transcripts represent mRNAs, which have links to p53 biology, but in general do not represent core p53 transcriptional targets. When combined with RNase-coupled proteomic analysis, which reveals RNA-dependent interactions between p53 and eukaryotic initiation factors and the 40S ribosome, we can conclude that p53 is binding protein coding mRNAs that are, or can be, translated. We can hypothesise that p53 can therefore regulate the translation of a specific subset of mRNAs, that represent a translational programme quite distinct from its transcriptional programmes. Further work is therefore required to establish the mechanisms

by which p53 is able to exert translational control, with iCLIP data, RNA-seq and ribosome profiling key in this effort. Stress-specific RNA binding is observed, but further work is required to elucidate the exact reasons behind this. A novel in silico method corroborates the potential of RNA binding at p53's DNA binding and C-terminal domains.

6

Global analysis of RBP responses to DNA damage

6.1 Introduction

RNA binding proteins' key role in RNA metabolism enables them to participate in the fine-tuning of gene expression at the post-transcriptional level. They are therefore crucial in the orchestration of cellular identity, development and response to environmental stimuli and cellular stressors. Since its advent, RIC has increasingly been applied to different cell types and organisms; highlighting cell-type specificities in RBPomes, and enabling identification of evolutionarily conserved RBPs [13]. RIC can also be utilised to evaluate the dynamic changes in a cells' RBP repertoire under different biological conditions. Vast remodelling of the RBPome has been observed upon viral infection [1], and treatment with arsenite [2]. An extension of RIC developed in our lab termed comparative RNA interactome capture (cRIC), therefore enables study of the dynamics of cellular RBPs under various conditions.

Maintenance of genome stability and the effective repair of DNA damage is crucial for cellular function and survival. Cells are subject to an estimated 10^5 spontaneous lesions per day from endogenous sources alone, with exogenous sources such as ionizing radiation (IR) and UV further contributing to cellular DNA damage [133]. Cells have therefore evolved a highly coordinated response system, termed the DNA damage response (DDR), with which they are able to sense DNA damage, and coordinate relevant downstream responses. Indeed, dysfunction or dysregulation of the DDR is prevalent in a range of cancers [55], with the accumulation of mutations and subsequent genome instability integral in carcinogenesis, and providing cancerous cells with selective advantage. Defects in genes encoding DDR pathway components are associated with cancer predisposition syndromes, such as BRCA-1/BRCA-2 syndrome [141]. Additionally, many chemotherapeutic agents and radiation therapy

regimens used to treat cancer rely on targeting of the DDR, causing DNA damage and subsequent cell death. As such, it is perhaps unsurprising that the p53 pathway plays a central role in the DDR, modulating late, downstream responses to DNA damage, such as cell cycle arrest, cellular senescence and apoptosis [55, 157].

Whilst the DDR has been extensively studied in terms of the roles of core DDR proteins, often termed sensors, mediators and effectors [140], emerging evidence has pointed to the importance of ncRNAs and RBPs in the DDR. Initially, it was thought that the involvement of RNAs and RBPs was purely through their ability to regulate the expression of other DDR components. However, recent accumulating evidence points to direct roles for ncRNAs and RBPs in the core DNA damage response [141, 158]. Evidence suggests the presence of RNAs at DNA damage sites, possibly as RNA:DNA hybrid R loops, which are increasingly attributed as a core component of the DDR [141]. This can then facilitate the recruitment of DNA repair factors to the damage foci [159, 133]. As techniques to identify RBPs have expanded, the number of DDR components identified as exhibiting additional RNA binding activities has expanded. Indeed, serial interactome capture of human lymphoblast nuclei reveals a network of core DNA damage response proteins, many of which are p53 pathway members, that exhibit RNA binding activities [160]. Elsewhere, 53BP1 was shown to bind non-poly(A) RNA in arsenite treated cells [2].

Metanalysis by Bader *et al.* [141] that looks at proteins that localise to damaged chromatin (specifically γH2AX, which acts as central marker in the DDR signalling pathway) and ssDNA, along with studies identifying proteins that are post translationally modified in response to DNA damage, reveal that 54% of proteins recruited to damaged chromatin, and 39% of proteins modified in response to damage, were

RBPs [141]. These works suggest a role for RBPs in mediating the DDR, through recruitment to damaged chromatin, and modification, in response to DNA damage. The regulation of DDR components is not sufficient to explain these associations between DNA damage scenarios and RBPs. Increasingly, RBPs are thought to play a significant, direct role in the DNA damage response.

Indeed, in etoposide treated GFP-p53 Hek293 cells, we observe a cluster of DNA damage associated factors coimmunoprecipitated with p53, that display RNase sensitivity (Chapter 5, figure 5.19). This includes characterised DDR associated RBPs XRCC6 (KU-70) and VCP, along with a number of core DDR proteins, that have previously been identified as p53 protein interaction partners, including 53BP1 (supporting previous findings [141]) along with multiple components of the MRE11/RAD50/NBN (MRN) complex. This supports the emerging ideas of the importance of RNA, and accordingly RBPs, in the DDR. Therefore, this chapter aims to utilise cRIC to identify RBPs involved in the DNA damage response.

6.2 Comparative RNA interactome capture allows comparison of RBPomes under different experimental conditions

RIC allows the identification of RNA binding proteins directly interacting with polyadenylated RNAs in living cells. A logical extension of the RIC protocol is its utilisation in examining the changes in cellular RBPomes using different treatment conditions; providing vital insights into the remodelling of cellular RBPomes under varying biological conditions. Our lab recently developed comparative RNA

interactome capture (cRIC), which enables the comparison of RIC datasets under varying cellular states. This was first utilised to examine the impact of Sindbis viral infection upon cellular RBPomes [1].

The cRIC protocol (Figure 6.1) was used to explore the impact of etoposide induced DNA damage on RNA binding protein dynamics in MCF-7 cells. Briefly, cRIC utilises stable isotope labelling by amino acids in cell culture (SILAC) [161]. SILAC involves culturing cells with light, medium or heavy amino acids, which are then incorporated into cellular proteins. MCF-7 cells were cultured with the relevant SILAC isotopes, and once 98% efficiency of label incorporation was confirmed, cells were treated with either a DMSO control (24 hours) or 10 μM etoposide. Two etoposide treatment durations were utilised, 6 hours or 24 hours, to examine temporal differences in the RBPome upon DNA damage. For each replicate, the SILAC isotopes were permutated across the 3 treatment conditions, to control for isotope-dependent effects. Cell monolayers were then UV irradiated to crosslink proteins and RNAs in direct contact, before cells were lysed, and inputs taken. At this point, the 3 samples comprising each replicate were mixed using equal quantities, further input aliquots taken, and oligo(dT) capture performed under denaturing conditions. RNase digestion enabled liberation of poly(A)-RNA bound proteins.

Resulting cRIC eluates were subject to SDS-PAGE followed by silver staining. Figure 6.2 shows silver stained gels for inputs and eluates obtained from the 5 cRIC replicates. This quality control step ensures reproducibility across the replicates. Eluates show clear differences in terms of protein composition compared to the inputs; representing dominant polyadenylated RNA bound proteins. Both input and eluates appear consistent across all 5 replicates. cRIC eluates were then analysed

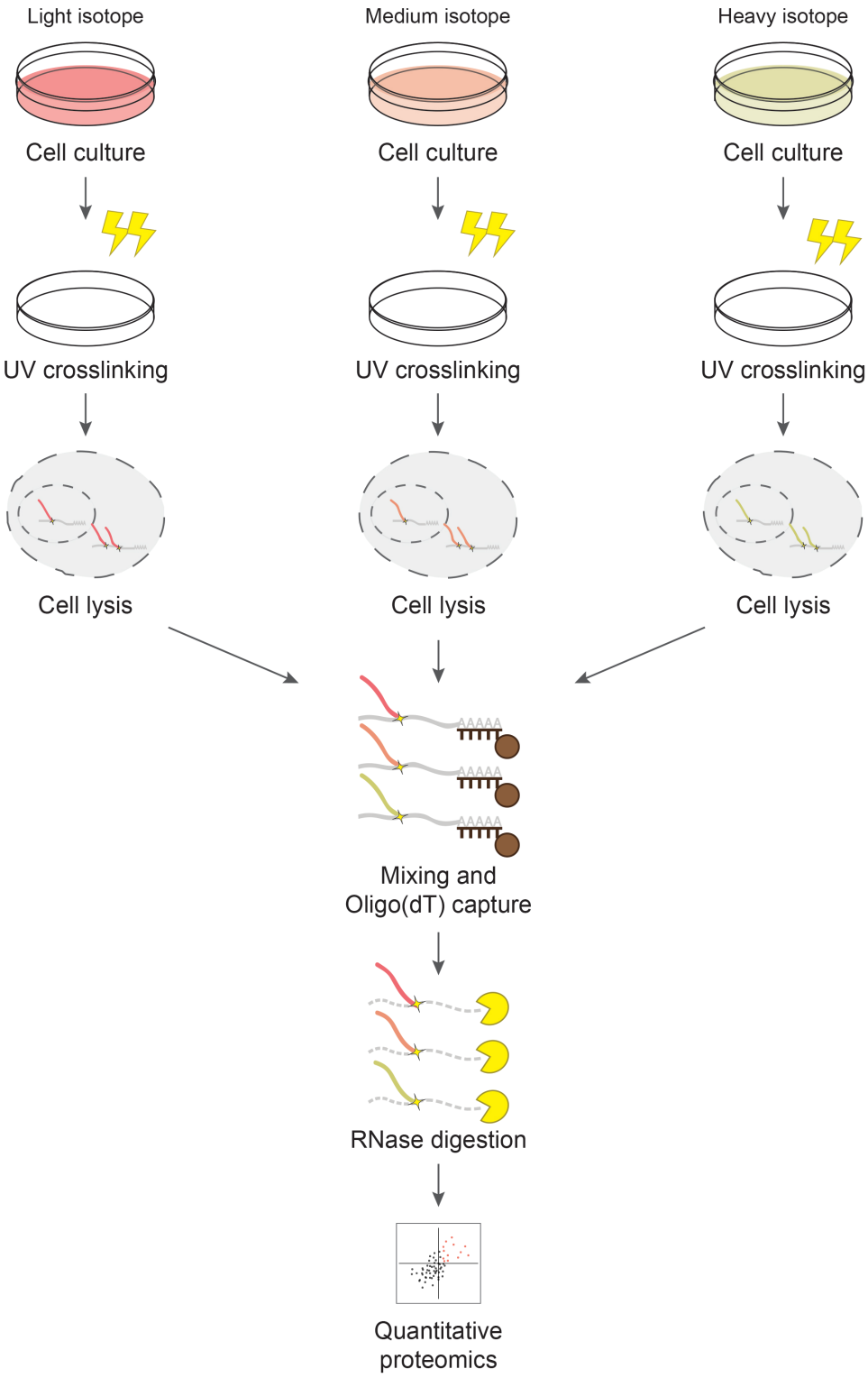


Figure 6.1: Comparative RNA interactome capture protocol. Schematic of the comparative RNA interactome capture (cRIC) protocol. Cells are grown in the presence of labelled amino acids, before treatment with 10 μM etoposide for 6 or 24 hours, or a DMSO control for 24 hours. SILAC labels are permutated across treatment conditions. UV irradiation induces covalent crosslinks between RNAs and proteins in direct contact. Cells are lysed, and lysates mixed in equal amounts. Oligo(dT) capture of poly(A) RNA is followed by stringent washing and elution of poly(A)-RBP complexes. Eluates are analysed by quantitative proteomics.

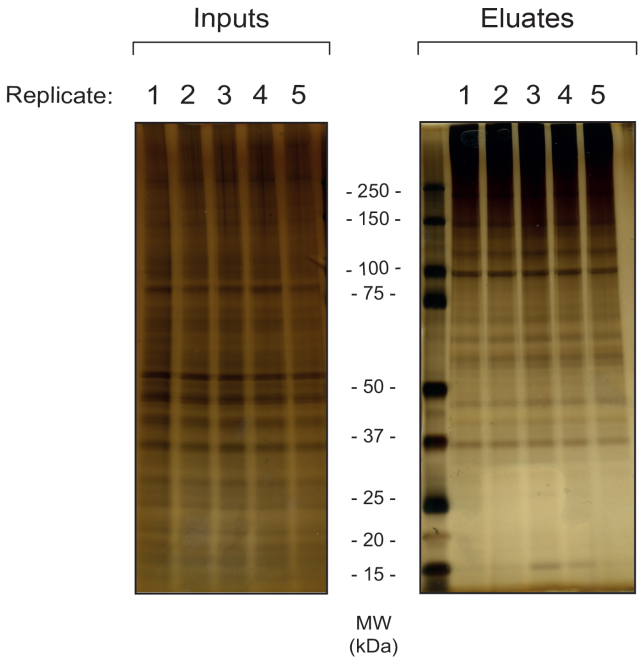


Figure 6.2: Silver staining of cRIC replicates. The cRIC protocol was performed on DMSO and etoposide treated, SILAC labelled MCF-7 cells, across 5 replicates. Inputs and eluates from each of the 5 cRIC replicates were subject to SDS-PAGE and silver staining.

by quantitative proteomics (performed by Svenja Hester).

6.3 Comparative RNA interactome capture yields high confidence RNA binding proteomes

A total of 916 proteins were identified across all replicates and experimental conditions. Of these, 715 proteins (78%) were previously identified human RBPs (Figure 6.3 A, based on Hentze *et al.* superset of human RBPs [11]), whilst 724 (79%) had GO annotation for 'RNA binding' (Figure 6.3 B). This is in keeping with previous RNA interactome experiments, and suggests acquisition of high confidence RNA binding proteins. This is corroborated by GO analysis (Figure 6.4), which reveals enrichment for terms associated with RNA metabolism.

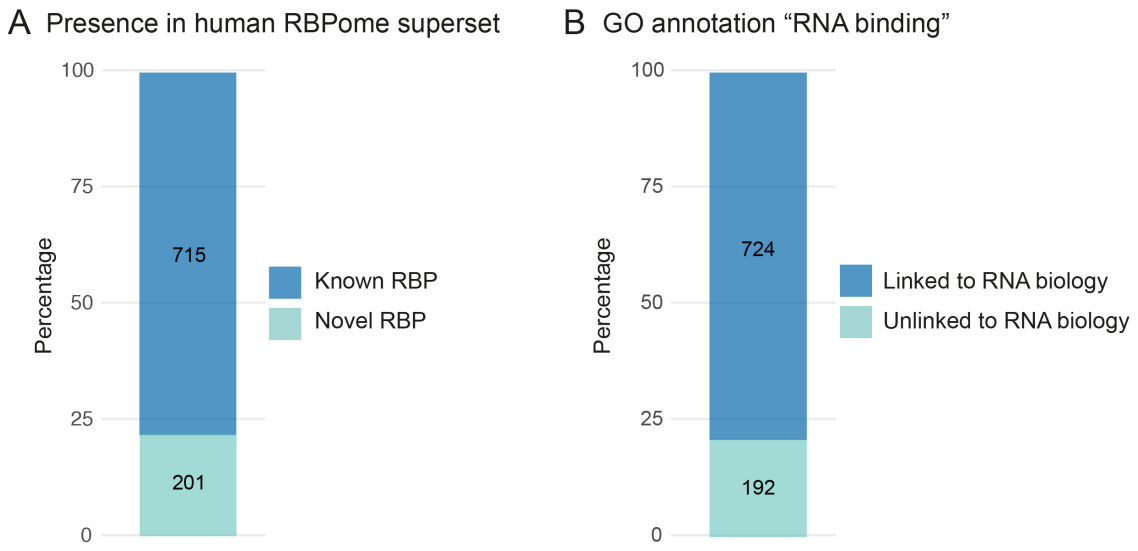


Figure 6.3: DNA damage cRIC identifies known RNA binding proteins Eluates obtained from DNA damage cRIC were analysed by quantitative proteomics. Proportion, and number, of proteins identified that A) are present in the *Hentze et al.* [11] compiled human RBP dataset, and B) are linked to RNA biology via GO annotation.

Categorisation of RNA binding domains, based on their experimental evidence, has lead to a tripartite classification system [14]. Classical RBDs are well characterised RBDs that have been experimentally validated in multiple systems, as defined by Lunde *et al* (2007) [162] , whilst nonclassical RBDs describe those domains shown to bind RNA in at least one experimental example, as in Castello *et al* (2012) [14]. Unknown RBDs are those with no prior experimental evidence of RNA binding. Of the 715 proteins identified, 223 contained classical RBDs (Figure 6.5 A), with the RNA recognition motif (RRM) by far the most prevalent RBD (Figure 6.5 B). Other well-defined RBDs including the DEAD box helicases, KH domains, zinc finger (ZF) domains, Piwi and PUF domains are also present in high numbers. 228 proteins contain nonclassical RBDs, with the WD40 domain the most prevalent nonclassical RBD (Figure 6.5 C). Other interesting nonclassical RBDs include lactate dehydrogenase 1 NAD binding and C-terminal domains (LDH1-N

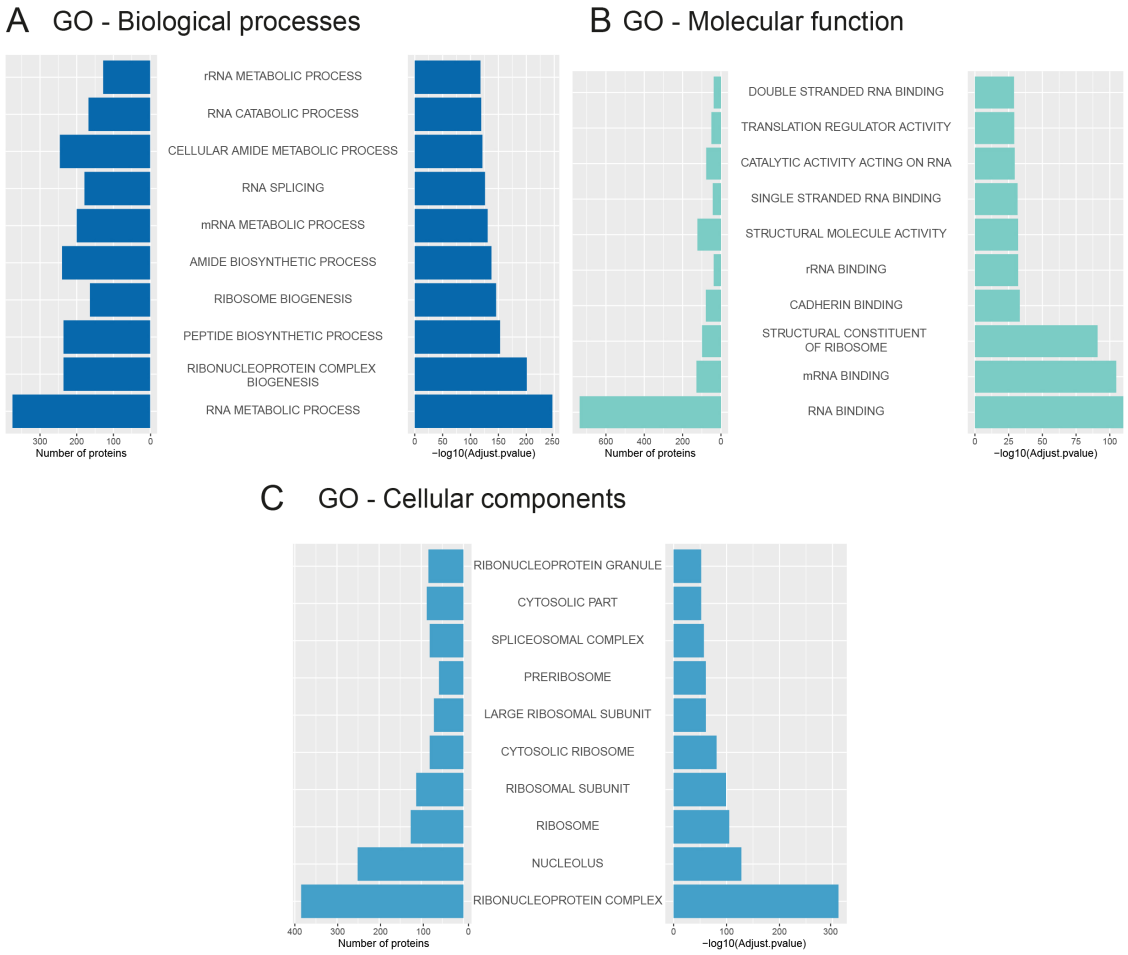


Figure 6.4: GO analysis of DNA damage cRIC identified proteins. GO analysis was performed on all proteins identified via DNA damage cRIC. Top 10 terms for A) biological processes, B) molecular function and C) cellular components are shown, according to adjusted P values.

and LDH1-C, respectively), along with the heat shock protein 70 (HSP70). These recapitulate previous works identifying heat shock proteins and enzymes of the intermediate metabolism as unorthodox RNA binding proteins that may bind via multifunctional domains, such as their enzymatic cores [8, 19]. Finally, there are 464 proteins (64.8%) lacking a previously defined RBD (Figure 6.5 A), in keeping with previous datasets where just over half of identified RBPs lack a discernable RBD [6]. Despite a larger number of proteins harbouring unknown RBDs, there are less clearly predominant domains represented, with a more even distribution of proteins across a wide range of domain architectures. Of the most abundantly represented (4 or 5 proteins per group), there are several interesting protein domains. The AAA 12 (ATPases associated with diverse cellular activities), Ras and Pkinase domains all represent catalytic protein domains, with the presence of such catalytic domains observed in previous RIC datasets [11]. The MA3, MIF4G and W2 domains have all been identified in eukaryotic initiation factors.

6.4 Comparative RNA interactome capture highlights novel DNA damage associated RBPs

A total of 137 proteins identified in this dataset had neither GO annotation, nor presence in the compiled human RBP dataset [11] (Figure 6.3). These potentially represent previously uncharacterised RBPs, henceforth termed candidate novel RBPs. GO analysis (Figure 6.6) demonstrates enrichment (false discovery rate < 0.05) for several immune response processes, along with several metabolism-related process, and a cluster of annotations associated with epithelial/epidermal cell layers (Figure

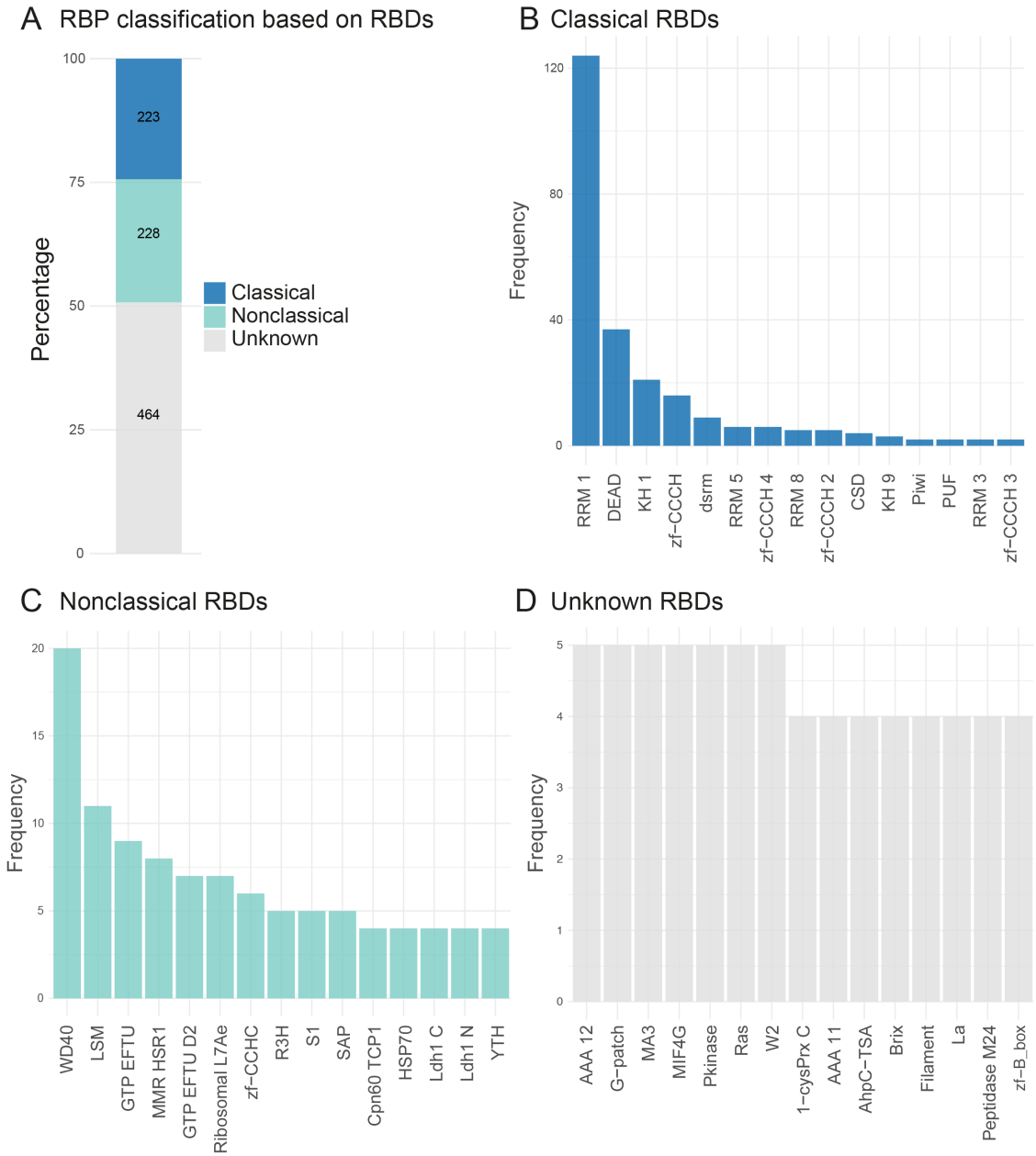


Figure 6.5: Domain architecture of DNA damage cRIC dataset. A) Number of proteins harbouring classical, non-classical or unknown RBDs in the DNA damage cRIC dataset. (B) Number of proteins annotated as possessing classical RBDs, (C) non-classical RBDs or (D) unknown RBDs.

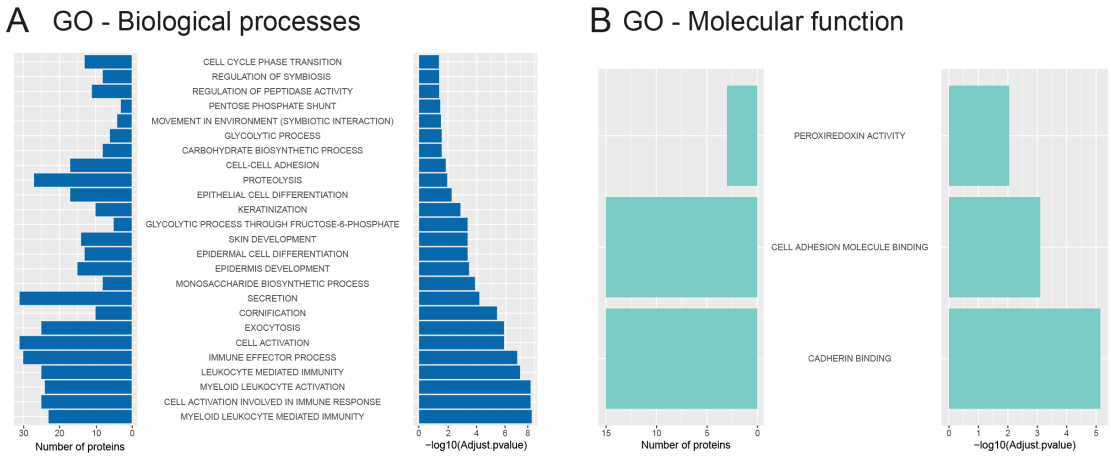


Figure 6.6: GO analysis of candidate novel RBPs identified by DNA damage cRIC. GO analysis was performed on candidate novel RBPs that had no previous link to RNA biology. Terms with an adusted P value < 0.05 are shown for A) biological processes, and B) molecular functions.

6.6 A). Interestingly, only 3 molecular function terms are statistically significant (Figure 6.6 B, FDR < 0.05), which suggest roles for these candidate novel RBPs in antioxidant metabolism ("peroxiredoxin activity"), and cell adhesion ("cell adhesion molecule binding" and "cadherin binding"). Noncanonical RBPs that exhibit enzymatic activity are a common feature of recent RIC experiments, whilst the role of proteins involved in cell adhesion requires further exploration.

6.5 Comparative RNA interactome capture reveals changes to the RBPome after DNA damage

To compare RBPomes obtained under the three experimental conditions, pairwise comparisons were performed between each of the treatment conditions, using moderated t-tests. For instances where there were missing values (an intensity of "zero"), and therefore infinite or zero ratios between the conditions, a semi-quantitative approach was used. Semi-quantitative candidates were selected based on presence in

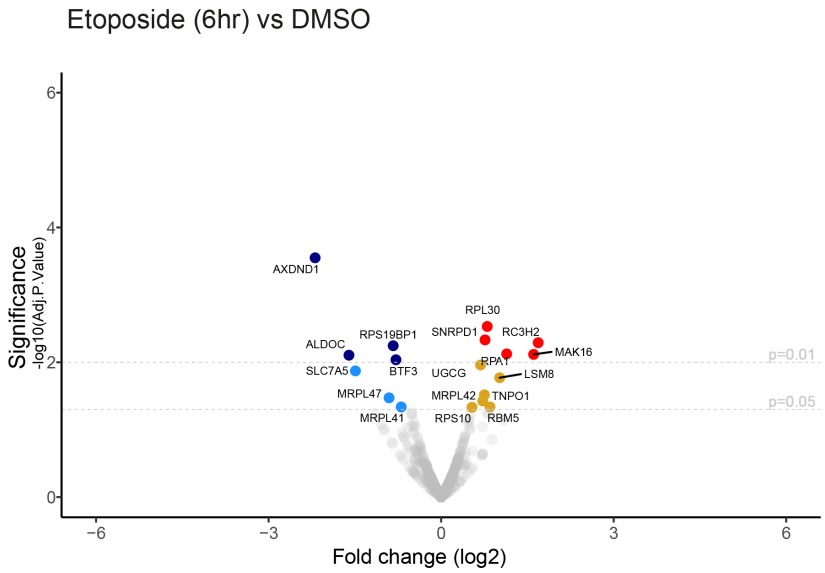


Figure 6.7: 6 hour etoposide cRIC. Volcano plot showing the log2 fold change and the significance (-log10 adjusted P value) of each protein between 6 hour etoposide treatment and DMSO control, using data from 5 biological replicates. Semi-quantitative candidates are not shown on the plot due to missing values, leading to zero/infinite ratios. These were selected based on presence in greater than 3 replicates for one condition, and less than 2 for the other; reflecting a potential bimodal on-off switch between treatment conditions.

greater than 3 replicates for one condition, and less than 2 for the other; reflecting a potential bimodal on-off switch between treatment conditions. Significant candidates were selected based on an adjusted P value of less than 0.05. Significant and semi-quantitative candidates for each condition were then pooled.

A total of 78 proteins exhibited differential interaction with RNA after 6 hour etoposide treatment, compared to control DMSO treatment. Of these, 17 showed increased association with RNA in etoposide treated cells, whilst 61 showed decreased RNA-association in etoposide treated cells (Figure 6.7). 49 of these 78 proteins were identified via the semi-quantitative method (not shown in Figure 6.7).

A total of 59 proteins exhibited differential interaction with RNA after 24 hour treatment with etoposide compared to control DMSO treatment. 14 showed increased association, whilst 45 showed decreased association with RNA in cells treated

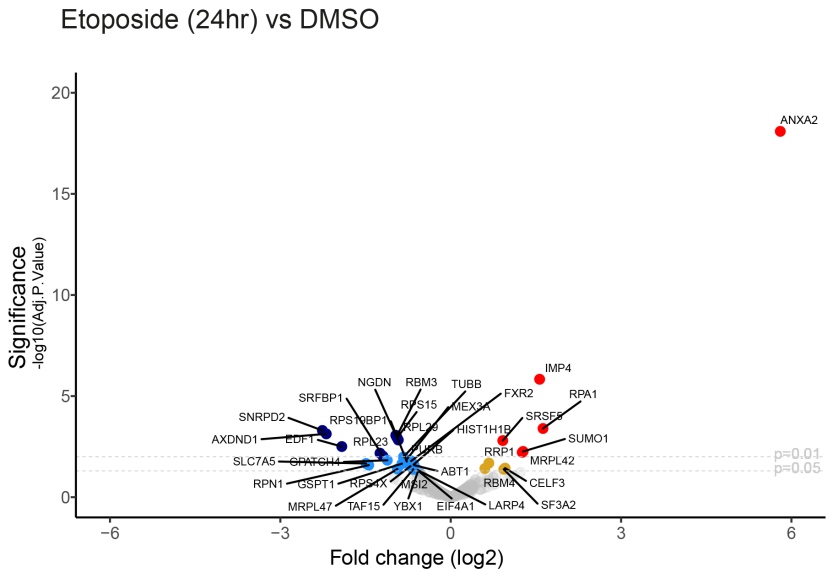


Figure 6.8: 24 hour etoposide cRIC. Volcano plot showing the log2 fold change and the significance (-log10 adjusted P value) of each protein between 24 hour etoposide treatment and DMSO control, using data from 5 biological replicates. Semi-quantitative candidates are not shown on the plot due to missing values, leading to zero/infinite ratios. These were selected based on presence in greater than 3 replicates for one condition, and less than 2 for the other; reflecting a potential bimodal on-off switch between treatment conditions.

with etoposide for 24 hours (Figure 6.8). 36 of these 59 proteins were identified via the semi-quantitative method (not shown in Figure 6.8).

When comparing the 6 hour and 24 hour etoposide treatments, a total of 89 proteins exhibited differential interaction with RNA. 6 hour treatment with etoposide resulted in the increased association of 52 proteins with RNA, whilst the 24 hour treated etoposide samples exhibited an increase in RNA association of 37 proteins (Figure 6.9). 35 of these 89 proteins were identified via the semi-quantitative method (not shown in Figure 6.9)

I then compared my cRIC dataset to a published dataset, in which mouse embryonic fibroblasts (MEFs) treated with either DMSO or etoposide (20uM, 6 hours) were subject to RIC and label free proteomic analysis [163]. Overall, my cRIC dataset identified almost 3 times as many poly(A) RNA bound proteins (916 compared to

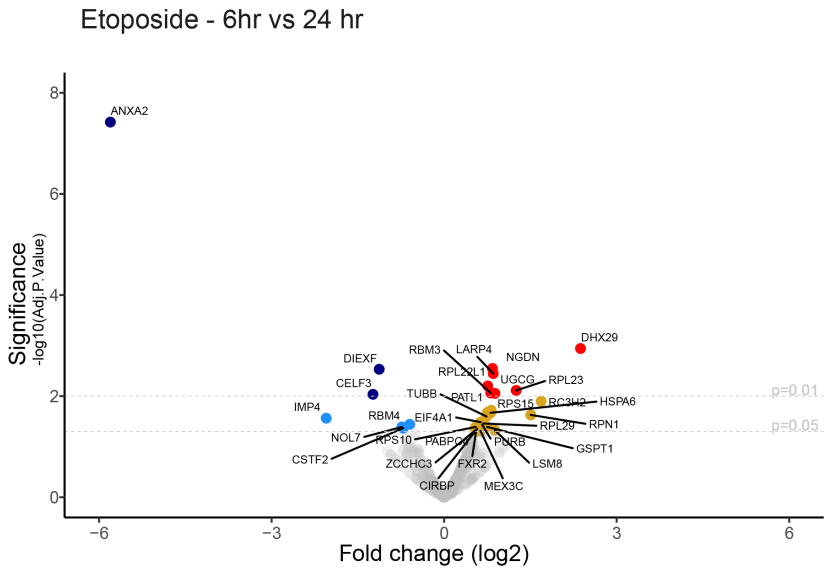


Figure 6.9: Comparing 6 and 24 hour etoposide treatments using cRIC. Volcano plot showing the log2 fold change and the significance (-log10 adjusted P value) of each protein between 6 hour and 24 hour etoposide treated cells, using data from 5 biological replicates. Semi-quantitative candidates are not shown on the plot due to missing values, leading to zero/infinite ratios. These were selected based on presence in greater than 3 replicates for one condition, and less than 2 for the other; reflecting a potential bimodal on-off switch between treatment conditions.

335, Figure 6.10). However, comparison of proteins exhibiting significantly altered RNA binding shows very little overlap between the two experiments (not shown). This may be largely due to the use of different cell lines. Cell-type has been proposed to play a significant role in differences in response to p53 stimulating agents, with cell-type specific chromatin architecture postulated as one of the main regulators of observed cell-type specific responses [147]. Additionally, different etoposide concentrations were used. Etoposide generates DNA damage in a dose-dependent manner [109], so varying concentrations may alter the downstream responses mounted by cells. Additionally, the different proteomic labelling and analysis approaches used may impact those proteins identified as significantly altered in comparison to DMSO controls.

To summarise, I have shown that cRIC can be implemented to study changes

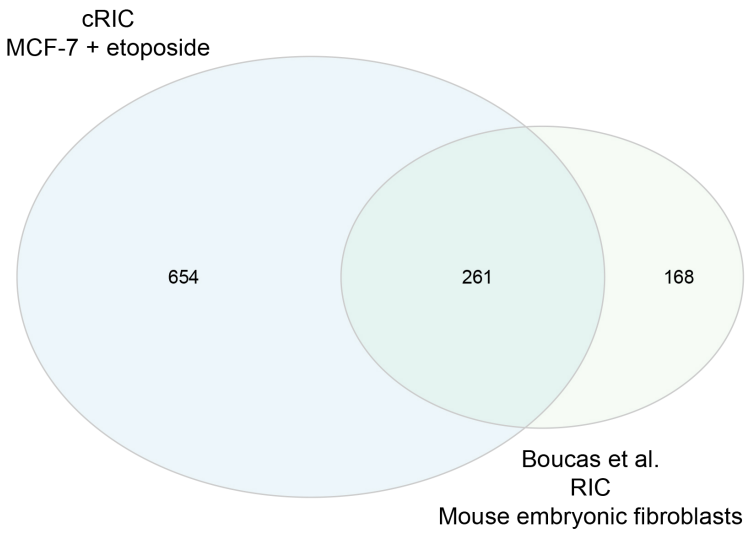


Figure 6.10: Comparison of DNA damage cRIC with etoposide RIC. Venn diagram showing the total number of proteins obtained in our DNA damage cRIC experiment (blue) compared to Boucas *et al.*'s etoposide RIC (green) [163], and the number of overlapping polyadenylated RNA bound proteins.

to cellular RBPomes, in order to study the DNA damage response; yielding RBPs whose RNA binding is significantly altered upon DNA damage. However, this is preliminary data, which requires further investigation.

6.6 Discussion

6.6.1 Comparative RNA interactome capture reveals DNA damage induced changes to the RBPome

Comparative RNA interactome capture was used to investigate how cellular RBPomes are remodelled upon etoposide induced DNA damage. Whilst extensive analysis of this dataset is ongoing, to date the data obtained provides proof of the utility of the cRIC protocol and analysis framework developed in our lab. In addition, the DNA damage cRIC dataset recapitulates results obtained in other RIC datasets. Firstly, the majority of proteins identified are either annotated with the GO "RNA binding"

term, or have been identified in previous RIC or RBDmap experiments, as defined by the Hentze *et al.* human RBP dataset [11]. Additionally, approximately half of the proteins identified do not possess a characterised RNA binding domain (either classical or nonclassical), which is in line with previous RIC results [11, 6]. Finally, this dataset identifies over 100 proteins that have previously not been annotated as RNA binding, or been experimentally identified as RBPs in the Hentze *et al.* human RBP dataset [11]; potentially further expanding the number of human proteins with RNA binding activities.

6.6.2 Future work

This dataset is very much at a preliminary stage, and therefore a significant amount of work is still to be performed. As with the standard RIC protocol, validatory western blots using both positive (such as PTBP1) and negative (such as B-actin) control antibodies need to be performed. DNA damage associated RBPs can then be identified, and validated by western blot. Combination of the cRIC dataset with RNA-seq and quantitative proteomic analysis of the whole proteome will reveal whether observed changes to RNA binding is due to altered target transcript abundance, changes in protein abundance, or changes in the intrinsic binding capacity of the proteins in question, by post translational modifications, for example [19, 7]. Once DNA damage responsive RBPs have been evaluated and validated, combination of our cRIC dataset with publically available eCLIP and iCLIP datasets will enable us to formulate post-transcriptional regulatory networks involved in the DDR.

Initially we probed this dataset for the presence of p53, but due to relatively

poor coverage, and the maintenance of p53 at low levels, even upon cellular stress, we were unable to identify p53 within this dataset. Analysis by western blot is more suited to investigation of specific proteins' binding to RNA in RIC experiments [10, 9], and enables identification of low abundance proteins. Alternatively, we can utilise pre-fractionation approaches to enable increased depth, and identification of much lower abundance proteins [164]. This will also assist with the high incidence of zero values, which was observed in the etoposide treated cRIC dataset, evident by the high number of proteins identified using the semi-quantitative approach.

Comparative RNA interactome capture therefore represents a hugely useful tool with which to study and compare RBPomes under varying cellular conditions. The application of cRIC to cells undergoing stressors associated with the cancer phenotype, and more specifically p53 signalling, such as UV irradiation, IR, hypoxia and nutrient deprivation, will help us further understand the role of post-transcriptional regulation in cancer associated pathways and processes.

7

Discussion

7.1 p53 binds RNA in living cells

The p53 tumour suppressor plays a crucial role in preventing neoplastic disease, through its mediation of responses to cellular stress. p53 is predominantly defined as a transcription factor, with modulation of gene expression key in its ability to mount stress responses. However, as the study of p53 has progressed over the past 40 years, the range of p53-implicated functions and processes has continued to expand.

The connection between p53 and RNA was first made through coelution of a C-terminal tryptic peptide of p53 with RNA in 1986 [64]. Since then, predominantly *in vitro* and acellular studies, utilising recombinant p53 proteins and *in vitro* transcribed RNAs, have characterised p53's propensity to bind RNAs. Barring a low stringency polynucleotide kinase assay performed by Yoshida *et al.* [74], evidence of p53-RNA interactions in living cells was still limited, and the biological function – if any – of such interactions, remained speculative.

Utilisation of RIC, which allows purification of poly(A) RNA bound proteins under denaturing conditions [10, 9], revealed p53 is able to bind RNA in living cells upon p53 stabilisation by nutlin, which was confirmed by orthogonal RNA labelling approaches. This provides proof of p53's ability to bind RNA in living cells, using highly stringent protein and RNA-centred approaches. When combined with the large number of RNA-dependent p53-coimmunoprecipitated RBPs, this provides strong evidence that p53 is a bona fide RNA binding protein.

Attempts to characterise p53's RNA binding partners in living cells reveals limitations in the application of conventional IgG antibodies to study p53-RNA interactions; evident in previous works [104], as well as my own. Presence of contaminant RNPs makes it impossible to confidently differentiate between direct and indirect

RNA interactions. Development of an inducible GFP-p53 system, combined with the high affinity and specificity GFP-trap nanobody, enables the utilisation of highly stringent purification procedures, which can be used to further study the role of p53's RNA binding activities.

7.2 RNA plays a role in p53 protein interactions

p53 has been shown to bind to a vast range of proteins interaction partners, which are crucial for regulation of p53 itself, as well as mediating p53's tumour suppressor functions [165]. Many of the high-throughput approaches used to identify potential interaction partners do not utilise nuclease treatment, and are prone to high levels of false positives [165]. Combination of RNase treatment with p53 immunoprecipitation in cells subject to DNA damage, revealed a large RNase sensitive p53 protein interaction network. This network represents proteins whose interactions with p53, in a proportion of cases, appears to involve an RNA bridge or intermediate. This does not discount direct protein interactions, but highlights a role for RNA-mediated interactions in some cases.

Interestingly, a cluster of DNA damage response factors are highlighted by this analysis, many of which are known p53 interaction partners. This contributes to the accumulating evidence that a number of core DDR components, many of which are components of the p53 pathway, exhibit RNA binding capacities [160, 17]. Additionally, screens for proteins involved in the DDR, and those modified in response to DNA damage, highlight RBPs as components of the conventional DDR [141]. In combination with work that reveals the presence, and necessity, of RNA species at DNA damage foci, this highlights the direct role of RBPs in DDRs. Utilisation

of comparative RNA interactome capture reveals significant remodelling of cellular RBPomes in response to etoposide induced DNA damage, but requires further analysis to understand the specific alterations that occur, and the role of identified RBPs in DNA damage responses.

7.3 p53 binds a distinct subset of mRNAs

Previous *in vitro* and acellular studies identified p53 as possessing non-specific RNA binding propensity associated with anti-helicase activities [72], along with examples of specific binding of p53 to a limited number of mRNA transcripts [80, 78, 79, 77]. Application of a CLIP-seq derived protocol confirms p53's interactions with a defined subset of mRNAs in cells, of which only 3 are core transcriptional targets of p53 [123]; suggesting p53's transcriptional and post-transcriptional programmes are distinct. In agreement, ribosome profiling in MCF-7 cells and BJ fibroblasts treated with nutlin [166], along with polysome profiling in serum starved HCT116 cell lines [167], reveal vastly different programmes of transcriptionally and translationally regulated genes. Indeed, Loayza-Puch *et al.* propose a bimodal tumour suppressive regulatory programme in response to p53 mediated senescence prompted by RAS activation [166], that exhibits selectivity regarding the mechanism of gene expression regulation (transcriptional or translational) conferred to different p53 responses [166].

RNA-dependent interactions of p53 with the 40S ribosome, eukaryotic initiation factors and PABPC1 suggest p53-bound RNAs are protein coding, translated mRNAs, thus linking p53's RNA binding activity with translational regulation. Indeed, recombinant p53 was shown to repress the translation of several mRNAs *in vitro*, with the primary mechanism thought to involve anti-helicase activity directed at

the 5'-UTR, thus preventing access of the translational machinery to the targetted RNAs. It is therefore reasonable to hypothesise that the p53-bound mRNA targets identified via CLIP-seq are susceptible to translational regulation by p53.

In combination, these works reveal a potential translational regulatory programme, that may contribute to p53-mediated responses. These works suggest that p53 may be able to regulate the translation of target genes, and that these represent a different subset of genes to those regulated by p53's conventional transcription factor activity. My CLIP-seq data fully supports these findings.

7.4 DNA and RNA binding proteins

An increasing number of DNA binding proteins are being identified as possessing RNA binding capabilities [17], termed DNA/RNA binding proteins or DRBPs. Up to 2% of the human proteome is estimated to comprise DRBPs [17]. As such, the concept that RNA binding proteins are functionally distinct from DNA binding proteins, and require study as a completely separate entity, is becoming decreasingly valid. Combination of DNA and RNA binding activities can be achieved in multiple ways. Binding of DNA and RNA concomitantly can represent a mechanism by which RNA 'cofactors' can regulate transcription. Alternatively, such simultaneous binding can enable deposition of RNA-containing complexes at specific loci [17]. Additionally, mutually exclusive binding of DNA and RNA enables regulation of transcription by competitive 'decoy' RNAs [17].

In the case of p53, *in vitro* studies have postulated that RNA and DNA binding are antagonistic, with binding of RNA to the C-terminus abrogating sequence specific binding at the core domain [74]. However, in their study of the *DINO* lncRNA,

Schmitt *et al.* demonstrate co-occupancy of *DINO* and p53 at p53-REs; suggesting concurrent binding of RNA and DNA, at least in this case [83]. How mutually exclusive binding would be regulated is unclear. Post translational modification status, subcellular localisation and interaction with *trans*-acting factors are all key modulators of RNA binding activities in RBPs, and could represent mechanisms by which binding substrate specificity is altered. For example, the regulated nucleocytoplasmic shuttling of polypyrimidine tract-binding protein (PTB), along with its protein interactions, dictate whether it functions as a splicing repressor or a translational activator [168]. Whether any of these factors are relevant in the modulation of p53-RNA interactions in living cells is as yet unclear. *In vitro* studies showed acetylation of the C-terminal domain at 4 lysine residues disabled RNA binding [76], suggesting, as with its transcription factor function, modulation of p53's RNA binding activity might be achieved, at least in part, by post translational modification. The way in which p53's RNA binding and DNA binding activities are balanced remains unclear, and warrants further study.

7.5 Post-transcriptional feedback loops

The p53 pathway is characterised by a number of autoregulatory loops, with the p53-MDM2 feedback loop providing the most obvious example. In these feedback loops, proteins whose activity or synthesis are regulated by p53 or p53 pathway members, alter p53 protein activity or stability. These feedback loops can be termed positive or negative autoregulatory feedback loops. Negative feedback loops play a role in homeostasis, enabling attenuation of the p53 response after activation by cellular stress, ensuring aberrant p53 signalling is not continued after the damage has been

repaired/removed [169]. Positive feedback loops, on the other hand play a role in stress responses and cell fate decisions; enabling amplification of stress signals into binary gene expression responses [169, 170, 171], which are crucial in p53-mediated responses to stress. Such autoregulatory loops also enable connection of the p53 pathway with other signal transduction pathways, which play key roles in responses to various stress stimuli [169].

Whilst initially these feedback loops involved p53 and downstream protein products, the presence and importance of feedback loops involving ncRNA species, including lncRNAs and miRNAs, is becoming apparent [150, 151, 84]. For example, the *DINO* lncRNA mentioned previously is a DNA damage induced p53 transcript, which then stabilises p53 protein; providing a positive feedback loop that amplifies the DNA damage signal once a threshold has been reached [83]. *Linc-ROR*, on the other hand, establishes a negative autoregulatory feedback loop with p53 by repressing p53 protein levels upon DNA damage [168]. As our understanding of the noncoding transcriptome expands, it is likely such examples will increase.

Identification of *p53, MDM2, MDM4* and *CUL7* transcripts as p53 targets in our CLIP-seq dataset suggest the presence of autoregulatory feedback loops also at the post-transcriptional level. Indeed, p53 has previously been shown *in vitro* to inhibit translation of its own RNA, through interaction with the 5'-UTR, which is then disrupted upon γ-irradiation [80]. Indeed, many well characterised RBPs exhibit such RBP-mediated autogenous feedback regulation [172]. Contrastingly, our dataset supports p53 binding to its own mRNA upon p53 activation. We can hypothesise that after p53 activation, p53 can bind to the mRNA of its own negative regulators, and impact their translation, generating either positive or negative post-

transcriptional regulatory feedback loops; contributing another layer to the already complex interplay between p53, its downstream targets, and its negative regulators. Ribosome profiling in GFP-p53 cells subject to genotoxic, non-genotoxic and control treatments will enable identification of transcripts whose translation is altered by p53, and in which direction. This can then be utilised to understand the impact of p53 on its negative regulators, and construct a post-transcriptional feedback model.

7.6 The RNA binding domain of p53

RNA binding capabilities have been attributed to both the core DNA binding domain and the regulatory C-terminal domain of p53, as predicted using RBDmap data. It is logical that the core DNA binding domain is able to bind other nucleic acid species, and not just DNA sequences comprising the p53-RE, due its innate nucleic acid binding characteristics. Conversely, p53's CTD is an appealing candidate for p53's RNA binding domain, as it mirrors the characteristics of a range of recently identified non-classical RBPs [7, 11, 19, 15]. The C-terminal regulatory domain is an intrinsically disordered region (IDR), which lacks stable tertiary structure in its native state [20]. Two other IDRs are present in p53, in the N-terminus, and the linker region, respectively, with each IDR having distinct amino acid compositions [20]. IDRs have relatively recently been identified as key RNA binding interfaces in a number of non-canonical RBPs, as well being able to combine in a modular manner with classical RBDs. Indeed, recent studies reveal that 20% of mammalian RBPs have $> 80\%$ disordered amino acids across their protein length [7]; highlighting the key role that IDRs can play in RNA binding. Intrinsic to IDRs is their ability to form adaptive binding conformations, facilitating interaction with a wide

range of interaction partners, which is evident in p53's high number of observed interaction partners [20]. IDRs are typically susceptible to extensive PTMs, as is the case for the C-terminus of p53 [52], which contains a high proportion of lysines and arginines that, while being PTM substrates, can also play a key role in mediating RNA binding activities [52]. IDRs in RNA binding domains show physiochemical similarities to analogous sequences in DNA binding proteins (DBPs), where these disordered basic tails act to promote linear diffusion across DNA, and effectively increase the capture radius of the DBPs as they search for binding sites (monkey bar theory) [173]. Indeed, this has been an observed function of the p53 C-terminal domain [173, 52]. A final feature of IDRs in RBPs, is their ability to promote phase transitions, which are central to the formation of stress granules, nuclear speckles and P-bodies. p53 has been identified in nuclear aggregates with polyadenylated RNA upon proteasomal inhibition [174], whilst cytoplasmic stress granules have been observed upon etoposide treatment [175]. As such, it is logical to suggest p53's intrinsically disordered regions may facilitate RNA granule formation; the functions of which are as yet unknown. The C-terminal domain is therefore an appealing candidate for p53's RNA binding domain, as this intrinsically disordered region reflects many of the characteristics of non-classical RBPs.

Even the most recent literature regarding p53's interactions with RNA reveal conflict regarding which domain is responsible for RNA binding [85, 83]. It may, of course, be possible that both core DNA binding and C-terminal domains are involved in forming contact with bound RNAs, in keeping with the cooperative modularity observed in classical RBDs. Indeed, it was previously shown that whilst the C-terminus is sufficient for RNA binding, the C-terminal fragment alone shows

decreased binding in comparison to full length protein [87, 76]; suggesting the importance of other sequences in forming RNA interactions.

Initial plans to characterise p53's RNA binding interface involved utilisation of the RBDmap protocol [15], with MCF-7 cells. RBDmap has already been used to identify the RNA binding domains of over 500 proteins, including non-canonical RBPs. RBDmap is able to identify IDRs involved in RNA binding, with over 170 RBPs identified where an IDR is the sole RNA binding domain [15]. As such, RBDmap would be able to verify whether the C-terminal IDR, or either of the other 2 p53 IDRs, is involved in RNA recognition. However, given the issues of depth observed in the cRIC experiment, it may be necessary to consider additional enrichment steps (e.g. IP) as well as the use of the Flp-In T-REx system to evaluate the impact of changes in p53 protein sequence on RNA binding.

7.7 Summary and future perspectives

To summarise, this work demonstrates p53 binds to poly(A) mRNAs in living cells. p53-bound target mRNAs represent a subset of targets that exhibit very little overlap with p53's transcriptional targets; suggesting p53's translational programme is distinct from its transcriptional regulatory programmes. In silico predictions of p53's RNA binding domains highlights regions of increased RNA binding probability within the core DNA binding and C-terminal domains, in keeping with *in vitro* works. RNase coupled proteomics reveals a network of RNA-sensitive p53 protein interactions, including a number of core DNA damage reponse components, which highlights the seeming importance of RNA in p53 mediated response to DNA damage.

Future works need to address the lack of clarity regarding the impact of p53 binding on target RNAs. Analysis of available iCLIP libraries will enable identification of p53's sequence and/or structural binding preferences. Utilisation of ribosome profiling and RNA-seq approaches under varying p53 activating conditions will enable validation of the hypothesised impact of p53 on mRNA translation, including the presence of post-transcriptional autorgegulatory feedback loops. Adaptation of the RBDmap protocol to include additional enrichment steps, along with the generation of p53 truncations and mutants in the Flp-In T-REx system, will enable identification of the protein regions comprising the p53 RNA binding domain. Finally, utilisation of comparative RNA interactome capture will enable identification of RNA binding proteins involved in the p53-mediated DNA damage response, as well as highlighting DNA damage response components that harbour previously unknown RNA binding activities.

Appendices

Supplementary table 1 – Common p53 coimmunoprecipitating proteins: proteins found to be coimmunoprecipitated in GFP-p53 IPs under both nutlin/doxycycline and etoposide/doxycycline treated cells.

Supplementary table 2 – CLIP-seq: Read counts for GFP normalised GFP-p53 + nutlin/doxycycline CLIP-seq samples. Only candidates present in 3/3 replicates are included.

Supplementary table 3 – CLIP-seq: Read counts for GFP normalised GFP-p53 + etoposide/doxycycline CLIP-seq samples. Only candidates present in 3/3 replicates are included.

Supplementary table 4 – cRIC: Proteins displaying differential RNA binding upon DNA damage.

Bibliography

(1) Garcia-Moreno, M., Noerenberg, M., Ni, S., Järvelin, A. I., González-Almela, E., Lenz, C. E., Bach-Pages, M., Cox, V., Avolio, R., Davis, T., et al. (2019). System-wide Profiling of RNA-Binding Proteins Uncovers Key Regulators of Virus Infection. *Molecular Cell 74*, 196–211.

(2) Trendel, J., Schwarzl, T., Horos, R., Prakash, A., Bateman, A., Hentze, M. W., and Krijgsveld, J. (2019). The Human RNA-Binding Proteome and Its Dynamics during Translational Arrest. *Cell 176*, 391–403.

(3) Esmaillie, R., Ignarski, M., Bohl, K., Krüger, T., Ahmad, D., Seufert, L., Schermer, B., Benzing, T., Müller, R. U., and Fabretti, F. (2019). Activation of Hypoxia-Inducible Factor Signaling Modulates the RNA Protein Interactome in Caenorhabditis elegans. *iScience 22*, 466–476.

(4) Milek, M., Imami, K., Mukherjee, N., De Bortoli, F., Zinnall, U., Hazapis, O., Trahan, C., Oeffinger, M., Heyd, F., Ohler, U., et al. (2017). DDX54 regulates transcriptome dynamics during DNA damage response. *Genome Research 27*, 1344–1359.

(5) Adriaens, C., Standaert, L., Barra, J., Latil, M., Verfaillie, A., Kalev, P., Boeckx, B., Wijnhoven, P. W., Radaelli, E., Vermi, W., et al. (2016). P53 induces formation of NEAT1 lncRNA-containing paraspeckles that modulate

replication stress response and chemosensitivity. *Nature Medicine 22*, 861–868.

(6) Beckmann, B. M., Castello, A., and Medenbach, J. (2016). The expanding universe of ribonucleoproteins: of novel RNA-binding proteins and unconventional interactions. *Pflügers Archiv European Journal of Physiology 468*, 1029–1040.

(7) Järvelin, A. I., Noerenberg, M., Davis, I., and Castello, A. (2016). The new (dis)order in RNA regulation. *Cell Communication and Signaling 14*, 1–22.

(8) Castello, A., Hentze, M. W., and Preiss, T. Metabolic Enzymes Enjoying New Partnerships as RNA-Binding Proteins, 2015.

(9) Castello, A., Horos, R., Strein, C., Fischer, B., Eichelbaum, K., Steinmetz, L. M., Krijgsveld, J., and Hentze, M. W. In *Methods in Molecular Biology*; Humana Press Inc.: 2015; Vol. 1358, pp 131–139.

(10) Castello, A., Horos, R., Strein, C., Fischer, B., Eichelbaum, K., Steinmetz, L. M., Krijgsveld, J., and Hentze, M. W. (2013). System-wide identification of RNA-binding proteins by interactome capture. *Nature Protocols 8*, 491–500.

(11) Hentze, M. W., Castello, A., Schwarzl, T., and Preiss, T. (2018). A brave new world of RNA-binding proteins. *Nature Reviews Molecular Cell Biology 19*, 327–341.

(12) Marondedze, C., Thomas, L., Serrano, N. L., Lilley, K. S., and Gehring, C. (2016). The RNA-binding protein repertoire of Arabidopsis thaliana. *Scientific Reports 6*, 1–13.

(13) Beckmann, B. M., Horos, R., Fischer, B., Castello, A., Eichelbaum, K., Alleaume, A. M., Schwarzl, T., Curk, T., Foehr, S., Huber, W., et al. (2015). The RNA-binding proteomes from yeast to man harbour conserved enigmRBPs. *Nature Communications 6*, 1–9.

(14) Castello, A., Fischer, B., Eichelbaum, K., Horos, R., Beckmann, B. M., Strein, C., Davey, N. E., Humphreys, D. T., Preiss, T., Steinmetz, L. M., et al. (2012). Insights into RNA Biology from an Atlas of Mammalian mRNA-Binding Proteins. *Cell 149*, 1393–1406.

(15) Castello, A., Frese, C. K., Fischer, B., Järvelin, A. I., Horos, R., Alleaume, A. M., Foehr, S., Curk, T., Krijgsveld, J., and Hentze, M. W. (2017). Identification of RNA-binding domains of RNA-binding proteins in cultured cells on a system-wide scale with RBDmap. *Nature Protocols 12*, DOI: `10.1038/nprot.2017.106`.

(16) Hentze, M. W., and Preiss, T. (2010). The REM phase of gene regulation. *Trends in Biochemical Sciences 35*, 423–426.

(17) Dutertre, M., and Vagner, S. (2017). DNA-Damage Response RNA-Binding Proteins (DDRBPs): Perspectives from a New Class of Proteins and Their RNA Targets. *Journal of Molecular Biology 429*, 3139–3145.

(18) Constable, A., Quick, S., Gray, N. K., and Hentze, M. W. (1992). Modulation of the RNA-binding activity of a regulatory protein by iron in vitro: Switching between enzymatic and genetic function? *Proceedings of the National Academy of Sciences of the United States of America 89*, 4554–4558.

(19) Balcerak, A., Trebinska-Stryjewska, A., Konopinski, R., Wakula, M., and Grzybowska, E. A. (2019). RNA–protein interactions: disorder, moonlighting and junk contribute to eukaryotic complexity. *Open Biology 9*, 1–13.

(20) Uversky, V. N. (2016). p53 proteoforms and intrinsic disorder: An illustration of the protein structure–function continuum concept. *International Journal of Molecular Sciences 17*, 1–37.

(21) Hanahan, D., and Weinberg, R. A. (2000). The Hallmarks of Cancer. *Cell 100*, 57–70.

(22) Hanahan, D., and Weinberg, R. A. (2011). Hallmarks of cancer: the next generation. *Cell 144*, 646–74.

(23) Kim, M. Y., Hur, J., and Jeong, S. (2009). Emerging roles of RNA and RNA-binding protein network in cancer cells. *BMB Reports 42*, 125–130.

(24) Kechavarzi, B., and Janga, S. C. (2014). Dissecting the expression landscape of RNA-binding proteins in human cancers. *Genome Biology 15*, 1–16.

(25) Pereira, B., Billaud, M., and Almeida, R. (2017). RNA-Binding Proteins in Cancer: Old Players and New Actors. *Trends in Cancer 3*, 506–528.

(26) Castello, A., Fischer, B., Hentze, M. W., and Preiss, T. (2013). RNA-binding proteins in Mendelian disease. *Trends in Genetics 29*, 318–327.

(27) Wurth, L., and Gebauer, F. (2015). RNA-binding proteins, multifaceted translational regulators in cancer. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms 1849*, 881–886.

(28) Dvinge, H., Kim, E., Abdel-Wahab, O., and Bradley, R. K. (2016). RNA splicing factors as oncoproteins and tumour suppressors. *Nature Reviews Cancer 16*, 413–430.

(29) El Marabti, E., and Younis, I. (2018). The cancer spliceome: Reprograming of alternative splicing in cancer. *Frontiers in Molecular Biosciences 5*, 1–11.

(30) Chen, W., Itoyama, T., and Chaganti, R. S. (2001). Splicing factor SRP20 is a novel partner of BCL6 in a t(3;6)(q27;p21) translocation in transformed follicular lymphoma. *Genes Chromosomes and Cancer 32*, 281–284.

(31) Lee, S. C. W., and Abdel-Wahab, O. (2016). Therapeutic targeting of splicing in cancer. *Nature Medicine 22*, 976–986.

(32) Yuan, F., Hankey, W., Wagner, E. J., Li, W., and Wang, Q. (2019). Alternative polyadenylation of mRNA and its role in cancer. *Genes & Diseases*, 1–12.

(33) Khabar, K. S. (2017). Hallmarks of cancer and AU-rich elements. *Wiley Interdisciplinary Reviews: RNA 8*, 1–25.

(34) Bhat, M., Robichaud, N., Hulea, L., Sonenberg, N., Pelletier, J., and Topisirovic, I. (2015). Targeting the translation machinery in cancer. *Nature Reviews Drug Discovery 14*, 261–278.

(35) Pelletier, J., Thomas, G., and Volarevic, S. (2018). Ribosome biogenesis in cancer: new players and therapeutic avenues. *Nature Publishing Group 18*, 51–63.

(36) Robichaud, N., Sonenberg, N., Ruggero, D., and Schneider, R. J. (2019). Translational control in cancer. *Cold Spring Harbor Perspectives in Biology 11*, 1–29.

(37) Truitt, M. L., and Ruggero, D. (2016). New frontiers in translational control of the cancer genome. *Nature Reviews Cancer 16*, 288–304.

(38) Sulima, S. O., Hofman, I. J., De Keersmaecker, K., and Dinman, J. D. (2017). How ribosomes translate cancer. *Cancer Discovery 7*, 1069–1087.

(39) Szostak, E., and Gebauer, F. (2013). Translational control by 3'-UTR-binding proteins. *Briefings in Functional Genomics 12*, 58–65.

(40) Durie, D., Lewis, S. M., Liwak, U., Kisilewicz, M., Gorospe, M., and Holcik, M. (2011). RNA-binding protein HuR mediates cytoprotection through stimulation of XIAP translation. *Oncogene 30*, 1460–1469.

(41) Moore, S., Järvelin, A. I., Davis, I., Bond, G. L., and Castello, A. (2018). Expanding horizons: new roles for non-canonical RNA-binding proteins in cancer. *Current Opinion in Genetics and Development 48*, 112–120.

(42) Forbes, S. A., Beare, D., Gunasekaran, P., Leung, K., Bindal, N., Boutselakis, H., Ding, M., Bamford, S., Cole, C., Ward, S., et al. (2015). COSMIC: Exploring the world's knowledge of somatic mutations in human cancer. *Nucleic Acids Research 43*, D805–D811.

(43) Amberger, J. S., Bocchini, C. A., Schiettecatte, F., Scott, A. F., and Hamosh, A. (2015). OMIM.org: Online Mendelian Inheritance in Man (OMIM®), an Online catalog of human genes and genetic disorders. *Nucleic Acids Research 43*, D789–D798.

(44) Blake, J. A., Christie, K. R., Dolan, M. E., Drabkin, H. J., Hill, D. P., Ni, L., Sitnikov, D., Burgess, S., Buza, T., Gresham, C., et al. (2015). Gene ontology consortium: Going forward. *Nucleic Acids Research 43*, D1049–D1056.

(45) Clucas, J., and Valderrama, F. (2014). ERM proteins in cancer progression. *Journal of Cell Science 127*, 267–275.

(46) Lianos, G. D., Alexiou, G. A., Mangano, A. A., Mangano, A. A., Rausei, S., Boni, L., Dionigi, G., and Roukos, D. H. (2015). The role of heat shock proteins in cancer. *Cancer Letters 360*, 114–118.

(47) Wang, X., Chen, M., Zhou, J., and Zhang, X. (2014). HSP27, 70 and 90, anti-apoptotic proteins, in clinical cancer therapy (review). *International Journal of Oncology 45*, 18–30.

(48) Lane, D. P., and Crawford, L. (1979). T antigen is bound to a host protein in SV 40-transformed cells. *Nature 278*, 261–263.

(49) Linzer, D. I., and Levine, A. J. (1979). Characterization of a 54K Dalton cellular SV40 tumor antigen present in SV40-transformed cells and uninfected embryonal carcinoma cells. *Cell 17*, 43–52.

(50) Brady, C. A., and Attardi, L. D. (2010). P53 At a Glance. *Journal of Cell Science 123*, 2527–2532.

(51) Levine, A. J., and Oren, M. (2010). The first 30 years of p53: growing ever more complex. *Nature Reviews Cancer 9*, 749–758.

(52) Laptenko, O., Tong, D. R., Manfredi, J., and Prives, C. (2016). The Tail That Wags the Dog: How the Disordered C-Terminal Domain Controls the

Transcriptional Activities of the p53 Tumor-Suppressor Protein. *Trends in Biochemical Sciences 41*, 1022–1034.

(53) Gaglia, G. Dynamics of p53 tetramers in live single cells, Ph.D. Thesis, 2014.

(54) Zhou, X., Cao, B., and Lu, H. (2017). Negative auto-regulators trap p53 in their web. *Journal of molecular cell biology 9*, 62–68.

(55) Meek, D. W., and Anderson, C. W. (2009). Posttranslational modification of p53: cooperative integrators of function. *Cold Spring Harbor perspectives in biology 1*, 1–17.

(56) Gu, B., and Zhu, W. G. (2012). Surf the Post-translational Modification Network of p53 Regulation. *International Journal of Biological Sciences 8*, 672–684.

(57) Zeron-Medina, J., Wang, X., Repapi, E., Campbell, M. R., Su, D., Castro-Giner, F., Davies, B., Peterse, E. F., Sacilotto, N., Walker, G. J., et al. (2013). A Polymorphic p53 Response Element in KIT Ligand Influences Cancer Risk and Has Undergo. *Cell 155*, 410–422.

(58) Kaiser, A. M., and Attardi, L. D. (2018). Deconstructing networks of p53-mediated tumor suppression in vivo. *Cell Death and Differentiation 25*, 93–103.

(59) Aubrey, B. J., Kelly, G. L., Janic, A., Herold, M. J., and Strasser, A. (2018). How does p53 induce apoptosis and how does this relate to p53-mediated tumour suppression? *Cell Death and Differentiation 25*, 104–113.

(60) Budanov, A. V. (2014). The role of tumor suppressor p53 in the antioxidant defense and metabolism. *Sub-Cellular Biochemistry 85*, 337–358.

(61) Hu, W., Zheng, T., and Wang, J. (2011). Regulation of fertility by the p53 family members. *Genes and Cancer 2*, 420–430.

(62) Baugh, E. H., Ke, H., Levine, A. J., Bonneau, R. A., and Chan, C. S. (2018). Why are there hotspot mutations in the TP53 gene in human cancers? *Cell Death and Differentiation 25*, 154–160.

(63) Sabapathy, K., and Lane, D. P. (2018). Therapeutic targeting of p53: All mutants are equal, but some mutants are more equal than others. *Nature Reviews Clinical Oncology 15*, 13–30.

(64) Samad, A., Anderson, C. W., and Carroll, R. B. (1986). Mapping of phosphomonoester and apparent phosphodiester bonds of the oncogene product p53 from simian virus 40-transformed 3T3 cells. *Proceedings of the National Academy of Sciences of the United States of America 83*, 897–901.

(65) Fontoura, B. M., Sorokina, E. A., David, E., and Carroll, R. B. (1992). p53 is covalently linked to 5.8S rRNA. *Molecular and Cellular Biology 12*, 5145–5151.

(66) Samad, A., and Carroll, R. B. (1991). The tumor suppressor p53 is bound to RNA by a stable covalent linkage. *Molecular and Cellular Biology 11*, 1598–1606.

(67) Fontoura, B. M. A., Atienza, C. A., Sorokina, E. A., Morimoto, T., and Carroll, R. B. (1997). Cytoplasmic p53 Polypeptide Is Associated with Ribosomes. *Molecular and Cellular Biology 17*, 3146–3154.

(68) Yin, X., Fontoura, B. M., Morimoto, T., and Carroll, R. B. (2003). Cytoplasmic complex of p53 and eEF2. *Journal of Cellular Physiology 196*, 474–482.

(69) Pipas, J. M., Ahuja, D., and Sa, M. T. (2005). SV40 large T antigen targets multiple cellular pathways to elicit cellular transformation. *Oncogene 24*, 7729–7745.

(70) Marechal, V., Elenbaas, B., Piette, J., Nicolas, J. C., and Levine, A. J. (1994). The ribosomal L5 protein is associated with mdm-2 and mdm-2-p53 complexes. *Molecular and Cellular Biology 14*, 7414–7420.

(71) Stürzbecher, H. W., Brain, R., Maimets, T., Addison, C., Rudge, K., and Jenkins, J. R. (1988). Mouse p53 blocks SV40 DNA replication in vitro and downregulates T antigen DNA helicase activity. *Oncogene 3*, 405–413.

(72) Oberosler, P., Hloch, P., Ramsperger, U., and Stahl, H. (1993). P53-Catalyzed Annealing of Complementary Single-Stranded Nucleic Acids. *The EMBO Journal 12*, 2389–2396.

(73) Nedbal, W., Frey, M., Willemann, B., Zentgraf, H., and Sczakiel, G. (1997). Mechanistic insights into p53-promoted RNA-RNA annealing. *Journal of Molecular Biology*, DOI: `10.1006/jmbi.1996.0813`.

(74) Yoshida, Y., Izumi, H., Torigoe, T., Ishiguchi, H., Yoshida, T., Itoh, H., and Kohno, K. (2004). Binding of RNA to p53 regulates its oligomerization and DNA-binding activity. *Oncogene 23*, 4371–4379.

(75) Riley, K. J., Cassiday, L. A., Kumar, A., and Maher, L. J. (2006). Recognition of RNA by the p53 tumor suppressor protein in the yeast three-hybrid system. *RNA 12*, 620–630.

(76) Riley, K. J. L., Ramirez-Alvarado, M., and Maher, L. J. (2007). RNA-p53 interactions in vitro. *Biochemistry 46*, 2480–2487.

(77) Galy, B., Créancier, L., Prado-Lourenço, L., Prats, A. C., and Prats, H. (2001). P53 directs conformational change and translation initiation blockade of human fibroblast growth factor 2 mRNA. *Oncogene 20*, 4613–4620.

(78) Miller, S. J., Suthiphongchai, T., Zambetti, G. P., and Ewen, M. E. (2000). p53 Binds Selectively to the 5' Untranslated Region of cdk4, an RNA Element Necessary and Sufficient for Transforming Growth Factor beta - and p53-Mediated Translational Inhibition of cdk4. *Molecular and Cellular Biology 20*, 8420–8431.

(79) Tournillon, A. S., López, I., Malbert-Colas, L., Findakly, S., Naski, N., Olivares-Illana, V., Karakostis, K., Vojtesek, B., Nylander, K., and Fåhraeus, R. (2017). P53 binds the mdmx mRNA and controls its translation. *Oncogene 36*, 723–730.

(80) Mosner, J., Bauer, C., and Deppert, W. (1995). Negative feedback regulation of wild-type p53 biosynthesis. *EMBO Journal 14*, 4442–4449.

(81) Lopez, I., Tournillon, A.-S., Nylander, K., and Fahraeus, R. (2015). p53-mediated control of gene expression via mRNA translation during Endoplasmic Reticulum stress. *Cell Cycle 14*, 3373–3378.

(82) Mlynarczyk, C., and Fåhraeus, R. (2014). Endoplasmic reticulum stress sensitizes cells to DNA damage-induced apoptosis through p53-dependent suppression of p21 CDKN1A. *Nature Communications 5*, 1–16.

(83) Schmitt, A. M., Garcia, J. T., Hung, T., Flynn, R. A., Shen, Y., Qu, K., Payumo, A. Y., Peres-Da-Silva, A., Broz, D. K., Baum, R., et al. (2016). An inducible long noncoding RNA amplifies DNA damage signaling. *Nature Genetics 48*, 1370–1376.

(84) Li, Y., and Lu, H. (2014). Noncoding RNAs: 'Our turn' to join the p53 network. *Journal of Molecular Cell Biology 6*, 179–180.

(85) Zhu, J., Liu, S., Ye, F., Shen, Y., Tie, Y., Zhu, J., Wei, L., Jin, Y., Fu, H., Wu, Y., and Zheng, X. (2015). Long noncoding RNA MEG3 interacts with p53 protein and regulates partial p53 target genes in hepatoma cells. *PLoS ONE 10*, 1–15.

(86) Yang, G., Lu, X., and Yuan, L. (2014). LncRNA: A link between RNA and cancer. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms 1839*, 1097–1109.

(87) Riley, K. J. L., and Maher, L. J. (2007). p53-RNA interactions: New clues in an old mystery. *RNA 13*, 1825–1833.

(88) Pfaffl, M. W., Horgan, G. W., and Dempfle, L. (2002). Relative expression software tool (REST) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. *Nucleic acids research 30*, e36.

(89) Hughes, C. S., Moggridge, S., Müller, T., Sorensen, P. H., Morin, G. B., and Krijgsveld, J. (2019). Single-pot, solid-phase-enhanced sample preparation for proteomics experiments. *Nature Protocols 14*, 68–85.

(90) Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature Biotechnology 26*, 1367–1372.

(91) Neuhauser, N., Michalski, A., Scheltema, R. A., Olsen, J. V., and Mann, M. (2011). Andromeda : A Peptide Search Engine Integrated into the MaxQuant Environment. *Journal of Proteome Research 10*, 1794–1805.

(92) Consortium, T. U. (2019). UniProt : a worldwide hub of protein knowledge. *Nucleic Acids Research 47*, 506–515.

(93) Zarnegar, B. J., Flynn, R. A., Shen, Y., Do, B. T., Chang, H. Y., and Khavari, P. A. (2016). IrCLIP platform for efficient characterization of protein-RNA interactions. *Nature Methods 13*, 489–492.

(94) Huppertz, I., Attig, J., Ambrogio, A. D., Easton, L. E., Sibley, C. R., Sugimoto, Y., Tajnik, M., König, J., and Ule, J. (2014). iCLIP : Protein – RNA interactions at nucleotide resolution. *Methods 65*, 274–287.

(95) König, J., Zarnack, K., Rot, G., Curk, T., Kayikci, M., Zupan, B., Turner, D. J., Luscombe, N. M., and Ule, J. (2010). ICLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. *Nature Structural and Molecular Biology 17*, 909–915.

(96) Van Nostrand, E. L., Pratt, G. A., Shishkin, A. A., Gelboin-Burkhart, C., Fang, M. Y., Sundararaman, B., Blue, S. M., Nguyen, T. B., Surka, C.,

Elkins, K., et al. (2016). Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). *Nature Methods 13*, 508–514.

(97) Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T. R. (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics 29*, 15–21.

(98) Quinlan, A. R. (2015). BEDTools: The Swiss-Army Tool for Genome Feature Analysis. *Current Protocols in Bioinformatics 47*, 1–11.

(99) Anders, S., Pyl, P. T., and Huber, W. (2015). HTSeq-A Python framework to work with high-throughput sequencing data. *Bioinformatics 31*, 166–169.

(100) Robinson, M. D., Mccarthy, D. J., and Smyth, G. K. (2010). edgeR : a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics 26*, 139–140.

(101) Mili, S., and Steitz, J. A. (2004). Evidence for reassociation of RNA-binding proteins after cell lysis: Implications for the interpretation of immunoprecipitation analyses. *RNA 10*, 1692–1694.

(102) Vassilev, L. T., Vu, B. T., Graves, B., Carvajal, D., Podlaski, F., Filipovic, Z., Kong, N., Kammlott, U., Lukacs, C., Klein, C., et al. (2004). In Vivo Activation of the p53 Pathway by Small-Molecule Antagonists of MDM2. *Science 303*, 844–848.

(103) Gupta, A., Shah, K., Oza, M. J., and Behl, T. (2019). Reactivation of p53 gene by MDM2 inhibitors: A novel therapy for cancer treatment. *Biomedicine and Pharmacotherapy 109*, 484–492.

(104) Riley, J., Kasandra J.L. & Maher (2007). Analysis of p53-RNA Interactions in Cultured Human Cells. *Biochemical and Biophysical Research Communications 363*, 381–387.

(105) Hamer-Casterman Atarchouch, C., T, Muyldermans, S., Robinson, G., Hamers, C., Bajyana, E., Bendahman, N., and Hamilton, R. (1993). Naturally occurring antibodies devoid of light chains. *Nature 363*, 446–448.

(106) Avolio, R., Järvelin, A. I., Mohammed, S., Agliarulo, I., Condelli, V., Zoppoli, P., Calice, G., Sarnataro, D., Bechara, E., Tartaglia, G. G., et al. (2018). Protein Syndesmos is a novel RNA-binding protein that regulates primary cilia formation. *Nucleic Acids Research 46*, 12067–12086.

(107) Hande, K. R. (1998). Etoposide: Four decades of development of a topoisomerase II inhibitor. *European Journal of Cancer 34*, 1514–1521.

(108) Montecucco, A., Zanetta, F., and Biamonti, G. (2015). Molecular mechanisms of etoposide. *EXCLI Journal 14*, 95–108.

(109) Wei, F., Hao, P., Zhang, X., Hu, H., Jiang, D., Yin, A., Wen, L., Zheng, L., He, J. Z., Mei, W., et al. (2018). Etoposide-induced DNA damage affects multiple cellular pathways in addition to DNA damage response. *Oncotarget 9*, 24122–24139.

(110) Karpinich, N. O., Tafani, M., Rothman, R. J., Russo, M. A., and Farber, J. L. (2002). The course of etoposide-induced apoptosis from damage to DNA and p53 activation to mitochondrial release of cytochrome c. *Journal of Biological Chemistry 277*, 16547–16552.

(111) Grandela, C., Pera, M. F., and Wolvetang, E. J. (2008). p53 is required for etoposide-induced apoptosis of human embryonic stem cells. *Stem Cell Research 1*, 116–128.

(112) Jamil, S., Lam, I., Majd, M., Tsai, S. H., and Duronio, V. (2015). Etoposide induces cell death via mitochondrial-dependent actions of p53. *Cancer Cell International 15*, 1–11.

(113) Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N. T., Morris, J. H., Bork, P., et al. (2019). STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Research 47*, D607–D613.

(114) Zhao, Z. The Regulation of Cellular Signalling by the 5S RNP, Ph.D. Thesis, 2018.

(115) Sloan, K. E., Bohnsack, M. T., and Watkins, N. J. (2013). The 5S RNP Couples p53 Homeostasis to Ribosome Biogenesis and Nucleolar Stress. *Cell Reports 5*, 237–247.

(116) Brannan, K. W., Jin, W., Huelga, S. C., Banks, C. A., Gilmore, J. M., Florens, L., Washburn, M. P., Van Nostrand, E. L., Pratt, G. A., Schwinn, M. K., et al. (2016). SONAR Discovers RNA-Binding Proteins from Analysis of Large-Scale Protein-Protein Interactomes. *Molecular Cell 64*, 282–293.

(117) Haberman, N., Huppertz, I., Attig, J., König, J., Wang, Z., Hauer, C., Hentze, M. W., Kulozik, A. E., Le Hir, H., Curk, T., et al. (2017). Insights

into the design and interpretation of iCLIP experiments. *Genome Biology 18*, 1–21.

(118) Ule, J., Jensen, K. B., Ruggiu, M., Mele, A., Ule, A., and Darnell, R. B. (2003). CLIP Identifies Nova-Regulated RNA Networks in the Brain. *Science 302*, 1212–1215.

(119) Zhao, Y., Zhang, Y., Teng, Y., Liu, K., Liu, Y., Li, W., and Wu, L. (2019). SpyCLIP: an easy-to-use and high-throughput compatible CLIP platform for the characterization of protein-RNA interactions with high accuracy. *Nucleic Acids Research 47*, 1–12.

(120) Lee, F. C., and Ule, J. (2018). Advances in CLIP Technologies for Studies of Protein-RNA Interactions. *Molecular Cell 69*, 354–369.

(121) Wheeler, E. C., Van Nostrand, E. L., and Yeo, G. W. (2018). Advances and challenges in the detection of transcriptome-wide protein–RNA interactions. *Wiley Interdisciplinary Reviews: RNA 9*, 1–11.

(122) Castelló, A., Izquierdo, J. M., Welnowska, E., and Carrasco, L. (2009). RNA nuclear export is blocked by poliovirus 2A protease and is concomitant with nucleoporin cleavage. *Journal of Cell Science 122*, 3799–3809.

(123) Fischer, M. (2017). Census and evaluation of p53 target genes. *Oncogene 36*, 3943–3956.

(124) Kuroda, T., Murayama, A., Katagiri, N., Ohta, Y. M., Fujita, E., Masumoto, H., Ema, M., Takahashi, S., Kimura, K., and Yanagisawa, J. (2011). RNA content in the nucleolus alters p53 acetylation via MYBBP1A. *EMBO Journal 30*, 1054–1066.

(125) Kumazawa, T., Nishimura, K., Katagiri, N., Hashimoto, S., Hayashi, Y., and Kimura, K. (2015). Gradual reduction in rRNA transcription triggers p53 acetylation and apoptosis via MYBBP1A. *Scientific Reports 5*, 1–13.

(126) Melotte, V., Qu, X., Ongenaert, M., van Criekinge, W., de Bruïne, A. P., Baldwin, H. S., and van Engeland, M. (2010). The N-myc downstream regulated gene (NDRG) family: diverse functions, multiple applications. *The FASEB Journal 24*, 4153–4166.

(127) Kovacevic, Z., Sivagurunathan, S., Mangs, H., Chikhani, S., Zhang, D., and Richardson, D. R. (2011). The metastasis suppressor, N-myc downstream regulated gene 1 (NDRG1), upregulates p21 via p53-independent mechanisms. *Carcinogenesis 32*, 732–740.

(128) Stein, S., Thomas, E. K., Herzog, B., Westfall, M. D., Rocheleau, J. V., Jackson, R. S., Wang, M., and Liang, P. (2004). NDRG1 is necessary for p53-dependent apoptosis. *Journal of Biological Chemistry 279*, 48930–48940.

(129) Croessmann, S., Wong, H. Y., Zabransky, D. J., Chu, D., Mendonca, J., Sharmaa, A., Mohseni, M., Rosen, D. M., Scharpf, R. B., Cidado, J., et al. (2015). NDRG1 links p53 with proliferation-mediated centrosome homeostasis and genome stability. *Proceedings of the National Academy of Sciences of the United States of America 112*, 11583–11588.

(130) Willmund, F., Del Alamo, M., Pechmann, S., Chen, T., Albanèse, V., Dammer, E. B., Peng, J., and Frydman, J. (2013). The cotranslational function of ribosome-associated Hsp70 in eukaryotic protein homeostasis. *Cell 152*, 196–209.

(131) Thorslund, T., Ripplinger, A., Hoffmann, S., Wild, T., Uckelmann, M., Villumsen, B., Narita, T., Sixma, T. K., Choudhary, C., Bekker-Jensen, S., and Mailand, N. (2015). Histone H1 couples initiation and amplification of ubiquitin signalling after DNA damage. *Nature 527*, 389–393.

(132) Li, Z., Li, Y., Tang, M., Peng, B., Lu, X., Yang, Q., Zhu, Q., Hou, T., Li, M., Liu, C., et al. (2018). Destabilization of linker histone H1.2 is essential for ATM activation and DNA damage repair. *Cell Research 28*, 756–770.

(133) Giono, L. E., Nieto Moreno, N., Cambindo Botto, A. E., Dujardin, G., Muñoz, M. J., and Kornblihtt, A. R. (2016). The RNA Response to DNA Damage. *Journal of Molecular Biology 428*, 2636–2651.

(134) Yang, K., Yang, J., and Yi, J. (2018). Nucleolar Stress: hallmarks, sensing mechanism and diseases. *Cell Stress 2*, 125–140.

(135) Bennetzen, M. V., Kosar, M., Bunkenborg, J., Payne, M. R., Bartkova, J., Lindström, M. S., Lukas, J., Andersen, J. S., Bartek, J., and Larsen, D. H. (2018). DNA damage-induced dynamic changes in abundance and cytosol-nuclear translocation of proteins involved in translational processes, metabolism, and autophagy. *Cell Cycle 17*, 2146–2163.

(136) Caudron-Herger, M., Rusin, S. F., Adamo, M. E., Seiler, J., Schmid, V., Barreau, E., Kettenbach, A. N., and Diederichs, S. (2019). R-DeeP: Proteome-wide and Quantitative Identification of RNA- dependent Proteins by Density Gradient Ultracentrifugation. *Molecular cell 75*, 184–199.

(137) Galy, B., Créancier, L., Zanibellato, C., Prats, A. C., and Prats, H. (2001). Tumour suppressor p53 inhibits human fibroblast growth factor 2 expression by a post-transcriptional mechanism. *Oncogene 20*, 1669–1677.

(138) Zaccara, S., Tebaldi, T., Pederiva, C., Ciribilli, Y., Bisio, A., and Inga, A. (2014). P53-directed translational control can shape and expand the universe of p53 target genes. *Cell Death and Differentiation 21*, 1522–1534.

(139) Moudry, P., Lukas, C., Macurek, L., Neumann, B., Heriche, J. K., Pepperkok, R., Ellenberg, J., Hodny, Z., Lukas, J., and Bartek, J. (2012). Nucleoporin NUP153 guards genome integrity by promoting nuclear import of 53BP1. *Cell Death and Differentiation 19*, 798–807.

(140) Su, M., Wang, H., Wang, W., Wang, Y., Ouyang, L., Pan, C., Xia, L., Cao, D., and Liao, Q. (2018). LncRNAs in DNA damage response and repair in cancer cells. *Acta Biochimica et Biophysica Sinica 50*, 433–439.

(141) Bader, A. S., Hawley, B. R., Wilczynska, A., and Bushell, M. (2020). The roles of RNA in DNA double-strand break repair. *British Journal of Cancer 122*, 613–623.

(142) D'Angelo, M. A. (2018). Nuclear pore complexes as hubs for gene regulation. *Nucleus 9*, 142–148.

(143) Singer, S., Zhao, R., Barsotti, A. M., Ouwehand, A., Fazollahi, M., Coutavas, E., Breuhahn, K., Neumann, O., Longerich, T., Pusterla, T., et al. (2012). Nuclear Pore Component Nup98 Is a Potential Tumor Suppressor and Regulates Posttranscriptional Expression of Select p53 Target Genes. *Molecular Cell 48*, 799–810.

(144) Ingolia, N. T. (2014). Ribosome profiling: New views of translation, from single codons to genome scale. *Nature Reviews Genetics 15*, 205–213.

(145) Arva, N. C., Talbott, K. E., Okoro, D. R., Brekman, A., Qiu, W. G., and Bargonetti, J. (2008). Disruption of the p53-Mdm2 complex by nutlin-3 reveals different cancer cell phenotypes. *Ethnicity and Disease 18*, 1–13.

(146) Yee-Lin, V., Pooi-Fong, W., and Soo-Beng, A. K. (2018). Nutlin-3, A p53-Mdm2 Antagonist for Nasopharyngeal Carcinoma Treatment. *Mini-Reviews in Medicinal Chemistry 18*, 173–183.

(147) Catizone, A. N., Good, C. R., Alexander, K. A., Berger, S. L., and Sammons, M. A. (2019). Comparison of genotoxic versus nongenotoxic stabilization of p53 provides insight into parallel stress-responsive transcriptional networks. *Cell Cycle 18*, 809–823.

(148) Shin, Y. J., Chen, K. Y., Sayed, A. H., Hencey, B., and Shen, X. (2013). Post-translational regulation enables robust p53 regulation. *BMC Systems Biology 7*, 1–12.

(149) Bode, A. M., and Dong, Z. (2004). Post-translational modification of p53 in tumorigenesis. *Nature Reviews Cancer 4*, 793–805.

(150) Chaudhary, R., and Lal, A. (2017). Long noncoding RNAs in the p53 network. *Wiley Interdisciplinary Reviews: RNA 8*, 1–17.

(151) Dangelmaier, E., Lazar, S. B., and Lal, A. (2019). Long noncoding RNAs: P53's secret weapon in the fight against cancer? *PLoS Biology 17*, 1–11.

(152) Huarte, M. (2016). P53 partners with RNA in the DNA damage response. *Nature Genetics 48*, 1298–1299.

(153) Uhl, M., Houwaart, T., Corrado, G., Wright, P. R., and Backofen, R. (2017). Computational analysis of CLIP-seq data. *Methods 118*, 60–72.

(154) Chakrabarti, A. M., Haberman, N., Praznik, A., Luscombe, N. M., and Ule, J. (2018). Data Science Issues in Studying Protein–RNA Interactions with CLIP Technologies. *Annual Review of Biomedical Data Science 1*, 235–261.

(155) Oughtred, R., Stark, C., Breitkreutz, B.-j., Rust, J., Boucher, L., Chang, C., Kolas, N., Donnell, L. O., Leung, G., Mcadam, R., et al. (2019). The BioGRID interaction database : 2019 update. *Nucleic Acids Research 47*, 529–541.

(156) Naski, N., Gajjar, M., Bourougaa, K., Malbert-Colas, L., Fåhraeus, R., and Candeias, M. M. (2009). The p53 mRNA-Mdm2 interaction. *Cell Cycle 8*, 31–34.

(157) Meek, D. W. (2004). The p53 response to DNA damage. *DNA Repair 3*, 1049–1056.

(158) Nishida, K., Kuwano, Y., Nishikawa, T., Masuda, K., and Rokutan, K. (2017). RNA binding proteins and genome integrity. *International Journal of Molecular Sciences 18*, 3–5.

(159) Crossley, M. P., Bocek, M., and Cimprich, K. A. (2019). R-Loops as Cellular Regulators and Genomic Threats. *Molecular Cell 73*, 398–411.

(160) Conrad, T., Albrecht, A. S., De Melo Costa, V. R., Sauer, S., Meierhofer, D., and Ørom, U. A. (2016). Serial interactome capture of the human cell nucleus. *Nature Communications 7*, 1–11.

(161) Ong, S.-e., Blagoev, B., Kratchmarova, I., Kristensen, D. B., Steen, H., Pandey, A., and Mann, M. (2002). Stable Isotope Labeling by Amino Acids in Cell Culture , SILAC , as a Simple and Accurate Approach to Expression Proteomics *. *Molecular & Cellular Proteomics 1*, 376–386.

(162) Lunde, B. M., Moore, C., and Varani, G. (2007). RNA-binding proteins: Modular design for efficient function. *Nature Reviews Molecular Cell Biology 8*, 479–490.

(163) Boucas, J., Fritz, C., Schmitt, A., Riabinska, A., Thelen, L., Peifer, M., Leeser, U., Nuernberg, P., Altmueller, J., Gaestel, M., et al. (2015). Label-free protein-RNA interactome analysis identifies khsrp signaling downstream of the p38/Mk2 kinase complex as a critical modulator of cell cycle progression. *PLoS ONE 10*, 1–19.

(164) Rockstroh, M., Müller, S., Jende, C., Kerzhner, A., Von Bergen, M., and Tomm, J. M. (2011). Cell fractionation - an important tool for compartment proteomics. *Journal of Integrated OMICS 1*, 135–143.

(165) Trinidad, A. G., Muller, P. A. J., Cuellar, J., Klejnot, M., Nobis, M., and Vousden, K. H. (2013). Interaction of p53 with the CCT Complex Promotes Protein Folding and Wild-Type p53 Activity. *Molecular Cell 50*, 805–817.

(166) Loayza-Puch, F., Agami, R., Drost, J., Rooijers, K., Lopes, R., Elkon, R., and Agami, R. (2013). p53 induces transcriptional and translational programs to suppress cell proliferation and growth. *Genome Biology 14*, 1–12.

(167) Liang, S., Bellato, H. M., Lorent, J., Lupinacci, F. C. S., Oertlin, C., van Hoef, V., Andrade, V. P., Roffé, M., Masvidal, L., Hajj, G. N. M., and

Larsson, O. (2018). Polysome-profiling in small tissue samples. *Nucleic Acids Research 46*, 1–13.

(168) Sawicka, K., Bushell, M., Spriggs, K. A., and Willis, A. E. (2008). Polypyrimidine-tract-binding protein: A multifunctional RNA-binding protein. *Biochemical Society Transactions 36*, 641–647.

(169) Harris, S. L., and Levine, A. J. (2005). The p53 pathway: Positive and negative feedback loops. *Oncogene 24*, 2899–2908.

(170) Lahav, G., Rosenfeld, N., Sigal, A., Geva-Zatorsky, N., Levine, A. J., Elowitz, M. B., and Alon, U. (2004). Dynamics of the p53-Mdm2 feedback loop in individual cells. *Nature Genetics 36*, 147–150.

(171) Wang, C., Xiang, L., Liu, H., and Zhou, J. (2019). Negative feedback and time delay regulate p53 oscillation in response to DNA damage. *Chinese Control Conference, CCC*, 6976–6981.

(172) Müller-McNicoll, M., Rossbach, O., Hui, J., and Medenbach, J. (2019). Auto-regulatory feedback by RNA-binding proteins. *Journal of Molecular Cell Biology 11*, 930–939.

(173) Vuzman, D., Azia, A., and Levy, Y. (2010). Searching DNA via a "Monkey Bar" Mechanism: The Significance of Disordered Tails. *Journal of Molecular Biology 396*, 674–684.

(174) Latonen, L., Moore, H. M., Bai, B., Jäämaa, S., and Laiho, M. (2011). Proteasome inhibitors induce nucleolar aggregation of proteasome target proteins and polyadenylated RNA by altering ubiquitin availability. *Oncogene 30*, 790–805.

(175) Díaz-muñoz, M. D., Turner, M., Kiselev, V. Y., and Novère, N. L. (2017). Tia1 dependent regulation of mRNA subcellular location and translation controls p53 expression in B cells. *Nature Communications 8*, 1–16.