The complex network of p53-regulated small non-coding RNAs and their gene targets in cancer

Jonathan Krell BSc (Hons) MBChB MRCP
MRC Clinical Research Fellow

Department of Surgery and Cancer
ICTEM Building, Hammersmith Hospital Campus
Imperial College, London, W12 0HS, UK

A Thesis submitted in accordance with the requirements of
Imperial College London for the degree of Doctor of Philosophy
February 2014

Supervisors: Professor Justin Stebbing & Dr. Leandro Castellano
Abstract

DNA damage transactivates TP53-regulated surveillance mechanisms that are crucial in maintaining cellular integrity and suppressing tumorigenesis. TP53 mediates this directly by transcriptionally modulating gene and microRNA (miRNA) expression and by regulating miRNA biogenesis through interaction with the DROSHA complex. However, the regulative mechanism of miRNA-AGO2 loading and the global change in AGO2 binding to its gene targets in response to DNA damage have not been investigated yet. In addition, the role of other non-coding RNAs, such as snoRNAs, in the TP53-mediated response to DNA damage has not yet been defined. Here we identify a novel group of TP53-regulated miRNAs and show that DNA damage induces and reduces the loading of a subset of miRNAs, including the let-7 family members onto AGO2, in a TP53-dependent manner and that this previously undescribed process is most likely the result of TP53 binding to AGO2. These findings indicate that TP53 control of AGO2 loading is a new mechanism of miRNA regulation in carcinogenesis. Using AGO2 RIP-Seq and PAR-CLIP we also show that TP53 modulates the reduction, induction and remodelling of AGO2 binding to the 3’UTR of different mRNA targets at specific RNA motifs. Furthermore, we determine on a transcriptome-wide level the miRNA-mRNA interaction networks involved in the response to DNA damage both in the presence or absence of TP53. We also show that those miRNAs whose cellular abundance or differential loading onto AGO2 is regulated by TP53, are involved in an intricate network of regulatory feedback and feedforward circuits that fine tune gene expression levels in response to DNA damage to permit DNA repair or the initiation of programmed cell death. Finally, we demonstrate a relationship between TP53 and the GAS5-derived snoRNAs both in cancer cell lines and human tissue samples which implies that this class of non-coding RNAs might also be involved in coordinating the TP53-mediated response. These findings provide a novel insight into the complexities surrounding the role of non-coding RNAs in the TP53 response to DNA damage and their relevance to carcinogenesis.
Statement of originality

All of the experimentation presented and written in this thesis has been conducted by me, apart from the following:

**Methods**    Laser capture microdissection (LCM) of colorectal cancer patient samples was performed by Carl Zeiss laboratories, Germany.

**Results**    Bioinformatic and network analyses were performed with the assistance of Dr Teresa Colombo, Sapienza University in Rome, Italy.

Jonathan Krell.

Copyright Declaration

The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives licence. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the licence terms of this work.
Acknowledgements

Foremost, I would like to thank both of my co-supervisors, Professor Justin Stebbing and Dr. Leandro Castellano for their excellent support and guidance throughout the duration of my PhD project. Leo thank you for your patience, technical advice, knowledge and ideas without which none of this would have been possible. Most importantly thanks for sharing the pain of the endless hours in the radioactive room optimising the PAR-CLIP experiment. Justin, thank you for giving me this opportunity, for sharing your unique ability to succeed and for your honesty, generosity and friendship over the past few years.

I am also indebted to the help, humour and friendship provided by my amazing lab mates Dr. Jimmy Jacob, Dr. Loredana Pellegrino, Laura Roca Alonso, Alex De Giorgio and Filipa Pinho. A special thanks to my friend and writing partner Mr. Adam Frampton and to Dr. Victoria Harding for keeping me balanced. Many thanks also to Dr. Teresa Colombo in Rome for her phenomenal support with our bioinformatic analyses. Thanks to Hua Zhang and Yichen Xu for their help with Western blotting and to Drs Laurence Game, Ivan Andrew, Nathalie Lambie and Keterina Rekopoulou in the MRC core laboratory for their expertise in next-generation sequencing. Many thanks also to Mr. Reza Mirnezami, Professor Jeremy Nicholson and Lord Ara Darzi for their collaboration in providing patient tissue samples. I would like to express my gratitude to Roger White for keeping us safe and to Shirley Line and Jeremy Batten for all of their administrative support. Furthermore, I am indebted to the Medical Research Council for funding my clinical research fellowship and to Mr. Anthony Hayman and all of the amazing people at Action Against Cancer for their indispensable financial contributions to our work— it simply would not have been possible without them. A special thanks also to Professors Mark Bower and Jonathan Waxman for sharing their humour, thoughts and advice and to Professors Charles Coombes and Bob Brown for creating an academically stimulating and exciting department in which to work.

I could not imagine that I would have been able to complete this PhD project without the encouragement of my family. I thank my wife, Vanessa, for putting up with me staying in the lab late at night and at weekends and for her love and support. I thank my beautiful son, Eden, for giving life another meaning and to my brothers Daniel and David for always being there. Finally, I would like to thank my mum and dad for inspiring me to try to make a difference.
Full manuscripts arising

Below are complete papers that have been published during this PhD period and those that are in submission. Only papers relating to non-coding RNAs have been included.


Index

Abstract .................................................................................................................................................. 1
Statement of originality ......................................................................................................................... 2
Copyright Declaration ............................................................................................................................ 2
Acknowledgements ................................................................................................................................. 3
Full manuscripts arising ........................................................................................................................... 4
Index ......................................................................................................................................................... 7
List of Figures .......................................................................................................................................... 11
List of Tables ........................................................................................................................................... 13
Abbreviations ......................................................................................................................................... 14

Chapter 1: Introduction ............................................................................................................................ 20
  1.1 MicroRNAs and their biogenesis ................................................................................................. 21
    1.1.1 AGO2 and miRNA loading on the RISC complex ........................................................ 25
    1.1.2 The let-7 family of miRNAs .......................................................................................... 27
  1.2 miRNAs and cancer ...................................................................................................................... 29
    1.2.1 MiRNAs as biomarkers ................................................................................................... 31
      1.2.1.1 Diagnostic biomarkers .......................................................................................... 31
      1.2.1.2 Prognostic biomarkers ........................................................................................ 33
      1.2.1.3 Predictive biomarkers ........................................................................................ 33
    1.2.2 MiRNA-modulating agents as cancer therapeutics .......................................................... 35
      1.2.2.1 Theory versus reality ........................................................................................... 35
      1.2.2.2 miRNA inhibition ................................................................................................. 36
      1.2.2.3 miRNA replacement/mimics ............................................................................... 38
  1.3 The importance of regulatory relationships between p53 and miRNAs in the p53-mediated cellular response to DNA damage ................................................................. 40
    1.3.1 p53 activation and its regulation of downstream signaling pathways following DNA damage ............................................................... 41
    1.3.2 The regulatory relationships between p53 and miRNAs ................................................... 45
      1.3.2.1 Transcriptional Regulation of miRNAs by p53 ....................................................... 46
      1.3.2.2 Post-transcriptional regulation of miRNA expression and function by p53 .......... 52
      1.3.2.3 The regulation of p53 expression by miRNAs ....................................................... 54
      1.3.2.4 Feedback loops involved in miRNA-regulated p53 signaling .............................. 55
1.3.3 Identifying p53-associated miRNA Targets .................................................................59
1.3.3.1 Cross-linking and immunoprecipitation (CLIP) .......................................................61
1.4 Small nucleolar RNAs ........................................................................................................65
1.4.1 The structure, synthesis and function of small nucleolar RNAs ....................................65
1.4.2 ‘Non-housekeeping’ roles of snoRNAs ...........................................................................68
1.4.3 SnoRNAs encoded for by the growth arrest-specific transcript 5 gene .......................72

Chapter 2: Materials and methods .......................................................................................74
2.1 Materials ............................................................................................................................75
2.1.1 Mammalian cell culture .................................................................................................75
2.1.2 Buffers and Solutions ....................................................................................................76
2.1.3 Antibodies/probes for western blots & immunoprecipitation ......................................78
2.1.4 Reagents used in the luciferase assays ..........................................................................79
2.1.5 Primers used for reverse transcription .........................................................................79
2.2 Methods .............................................................................................................................79
2.2.1 Mammalian cell culture ................................................................................................79
2.2.1a Growing and passaging cells ......................................................................................79
2.2.1b Cell treatments with doxorubicin .............................................................................80
2.2.2 Quantitative real-time Reverse Transcription-PCR ....................................................80
2.2.2a RNA preparation ..........................................................................................................80
2.2.2b cDNA synthesis ...........................................................................................................81
2.2.2c Quantitative real-time PCR ........................................................................................81
2.2.3 SDS-polyacrylamide gel electrophoresis and western blot .........................................81
2.2.3a Protein extraction ........................................................................................................81
2.2.3b Protein quantification ................................................................................................82
2.2.3c SDS-polyacrylamide gel electrophoresis ....................................................................82
2.2.3d Western blots .............................................................................................................82
2.2.4 Immunoprecipitation .....................................................................................................83
2.2.4a Preparation and washing of the sepharose beads .......................................................83
2.2.4b Preparation of the cell lysate ......................................................................................83
2.2.4c Conjugation of antibodies to the beads, and the immunoprecipitation steps ............83
2.2.4d Extraction of RNA from the immunoprecipitated samples using phenol chloroform ....................................................................................................................84
2.2.5 PARCLIP ......................................................................................................................84
2.2.6 Preparation of the RNA and small RNA libraries ......................................................87
2.2.7 3’-UTR luciferase reported assays .................................................................87
2.2.8 Collection, handling and RNA extraction from laser captured micro-dissected (LCM) tumour samples ..........................................................88
2.2.9 Collection, handling and RNA extraction from macro-dissected colorectal tissue 91
2.2.10 Next generation sequencing and bioinformatic analysis of results ..........91
   2.2.10a Description of samples and experimental design .................................91
   2.2.10b Processing of small RNA sequencing data ..........................................92
   2.2.10c Analysis of miRNA modulation following DNA-damage .................95
   2.2.10d Processing of long RNA sequencing data ..........................................96
   2.2.10e Seed enrichment analysis .................................................................103
   2.2.10f Processing of AGO2 PAR-CLIP data ...............................................103

Chapter 3: Results .................................................................................................. 108

3.1 The post-transcriptional regulation of gene expression by p53 is mediated by its ability to regulate miRNA abundance, control miRNA loading onto AGO2 and remodel the miRNA-mRNA interaction network through novel mechanisms .........................109
   3.1.1 A small RNA-seq approach identifies novel miRNAs that are transcriptionally regulated following DNA damage in a p53-dependent and –independent manner ....112
   3.1.2 A small RIP-seq approach demonstrates that DNA damage induces a previously undescribed mechanism that controls miRNA function post-transcriptionally by regulating the loading of a subset of miRNAs onto AGO2 in a p53-dependent manner ..............................................................117
   3.1.3 AGO2 loading of certain miRNAs, including the let-7 family, is regulated in a p53-dependent manner following DNA damage and this process is selective to those cell lines expressing wild-type rather than mutated p53..............................118
   3.1.4 Increased loading of the let-7 family onto AGO2 following DOX-induced DNA damage is not related to an early transcriptional effect ........................................120
   3.1.5 The differential loading of identified miRNAs onto AGO2 following DOX-induced DNA damage is not due to changes in AGO2 expression .................................120
   3.1.6 Combining our small RIP-seq and small RNA-seq analyses validates a group of miRNAs that are regulated in a p53-dependent or p53-independent manner following DNA damage ........................................................................121
   3.1.7 An RNA-seq approach followed by an extensive bioinformatic analysis identifies genes whose abundance is regulated in a p53-dependent manner following DNA damage and illustrates their function within cellular signaling networks ........................................122
   3.1.8 An RNA-seq approach followed by an in depth bioinformatic analysis identifies genes whose abundance is regulated in a p53-independent manner following DNA damage and illustrates their function within cellular signaling networks ..........124
3.1.9 p53 post-transcriptionally modulates the activity of a subgroup of genes involved in the DNA damage response through differential targeting by AGO2 and this process involves p53-regulated miRNAs ................................................................. 128

3.1.10 An AGO2 PARCLIP analysis illustrates the p53-associated miRNA/mRNA interactome at a genome wide level and establishes the functional effects of miRNA/mRNA binding in terms of p53-regulated signaling networks ...................... 132

3.1.11 Luciferase reporter assays confirm that CCND1 and POGZ are targeted by let-7a, let-7d, miR-23a and miR-34a and validate our PARCLIP approach for defining targets of miRNAs regulated upon DNA damage ........................................ 146

3.1.12 p53 regulates the activity of a subgroup of genes involved in the DNA damage response by modulating AGO2 binding on their UTR sequences ........................................ 147

3.2 The GAS5-derived snoRNAs are involved in the p53-dependent cellular response to DNA damage ................................................................................................................. 151

3.2.1 Doxorubicin-induced DNA damage increases GAS5-derived snoRNA expression in a p53 dependent manner in colorectal cancer cell lines ......................... 152

3.2.2 GAS5-derived snoRNA expression varies between normal and malignant colorectal fresh non-microdissected tissue in a p53-dependent manner .................. 155

3.2.3 GAS5-derived snoRNA expression varies between normal, pre-malignant and malignant microdissected FFPE colorectal tissue and levels correlate with miR-34a expression ........................................................................................................... 159

3.2.4 The expression of GAS5-derived snoRNAs is not affected in colorectal cancer cell lines in which DICER has been knocked-down and therefore do not appear to be processed by DICER ......................................................................................... 161

Chapter 4: Discussion ............................................................................................................. 170

4.1 The p53-regulated small RNA network and its role in the DNA damage response. 171

4.1.1 We identified a novel group of miRNAs regulated by DNA damage and found that this stress stimulus induced the differential loading of specific miRNAs onto AGO2, although the mechanism responsible requires further evaluation .............. 171

4.1.2 How does p53 mediate the remodelling of AGO2 binding to the 3UTR of different targets at specific RNA motifs? ................................................................. 176

4.1.3 Feedback and feed-forward loops provide the means for p53 to survey and coordinate the cellular response to DNA damage and complex miRNA:mRNA regulatory networks enable fine-tuning of this response .................................... 178

4.1.4 DNA damage induces the expression of the GAS5-derived snoRNAs in a p53-dependent manner in colorectal cancer cell lines, and a positive correlation exists between p53 and the same snoRNAs in normal, pre-malignant and malignant colorectal tissue ........................................................................................................ 182

4.1.5 The potential role of p53-regulated non-coding RNAs in cancer therapeutics...... 185

4.1.6 Future Directions ........................................................................................................... 186
List of Figures

Figure 1: miRNA processing. ................................................................. 25
Figure 2: Examples of p53-regulated feedback/feed-forward loops involving miRNAs. ............ 56
Figure 3: Coherent and incoherent feed-forward loops ......................................................... 57
Figure 4: Diagrammatic representation of an AGO PAR-CLIP experiment ................................. 64
Figure 5: Photographs of tissue blocks on microdissection slides at various stages of the procedure. 89
Figure 6: Outline of small RNA sequencing data processing .................................................. 93
Figure 7: Unsupervised clustering of miRNA samples based on normalized read counts. ............ 94
Figure 8: Histograms demonstrating miRNA mean fold changes in doxorubicin-treated vs. untreated samples................................................................. 97
Figure 9: Scatterplots of Log Fold-Change vs Average Expression for TP53+/+ cells .................. 98
Figure 10: Scatterplots of Log Fold-Change vs Average Expression for TP53-/- cells .................. 99
Figure 11: Outline of long RNA sequencing data processing .................................................. 100
Figure 12: Histograms of genes mean fold changes in doxorubicin-treated vs. untreated samples... 101
Figure 13: Scatterplots of genes mean fold changes in doxorubicin-treated vs. untreated samples as a function of mean gene abundance......................................................... 102
Figure 14: Distribution of mismatches to the human genome for PAR-CLIP mapped reads .......... 104
Figure 15: Outline of PAR-CLIP data processing ................................................................. 105
Figure 16: Summary of the cell treatments, sample processing and sequencing of samples used .... 111
Figure 17: A combined RNA-seq and RIP-seq approach identifies miRNAs showing differential expression and/or AGO2 loading following DOX-induced DNA damage ............................................. 114
Figure 18: A combined small RNA-seq and AGO2-RIP seq approach confirms that the loading of members of the let-7 family onto AGO2 is regulated in a p53-dependent manner following DOX-induced DNA damage ................................................................................. 115
Figure 19: RT-qPCR and Western blotting validate our small RNA-seq and RIP-seq analyses and demonstrate that the AGO2 loading affect is dependent on the presence of wild-type p53 and is not the result of changes in AGO2 abundance .................................................................................. 116
Figure 20: Heat map illustrating the differential expression of a number of genes relevant to the DNA damage response that were found to be differentially expressed between DOX and vehicle treated p53WT and p53null HCT116 cells in either the total RNA or AGO IP samples .................................................. 123
Figure 21: Pearson correlation plots demonstrate consistency between replicates in terms of gene expression changes in the RNA-seq experiments .............................................................................. 124
Figure 22: Top GO-term and KEGG analyses of genes identified as being up-regulated in a p53-dependent manner following DOX treatment in the RNA-seq.................................................. 125
Figure 23: Top GO-term and KEGG analyses of genes identified as being down-regulated in a p53-dependent manner following DOX treatment in the RNA-seq .................................................. 126
Figure 24: Top GO-term and KEGG analyses of genes identified as being up-regulated in a p53-independent manner following DOX treatment in the RNA-seq ......................................................................... 127
Figure 25: Top GO-term and KEGG analyses of genes identified as being down-regulated in a p53-independent manner following DOX treatment in the RNA-seq ......................................................................... 128
Figure 26: Many AGO2-enriched genes, including RET, CDK6 and FGF18 are targets of the p53-regulated miRNAs, including in some cases those miRNAs that were more loaded onto AGO2 following DOX treatment and these genes are involved in many cellular signalling pathways that p53 would seek to inhibit after DNA damage..........................................................................................130
Figure 27: A seed enrichment analysis strongly predicts that AGO2-enriched genes are targets of p53-regulated miRNAs ........................................................................................................................................131
Figure 28: A DAVID analysis of the RIP-seq genes .................................................................................................................................132
Figure 29: Cytoscape illustrations depicting miRNA:mRNA interaction networks involved in the DNA damage response as identified by a PAR-CLIP analysis ..................................................................................................................136
Figure 30: Top GO-term and KEGG analyses of genes identified by PAR-CLIP to be targeted by miRNAs that were upregulated in a p53-dependent manner in the RNA-seq .......................................................137
Figure 31: Top GO-term analyses of genes identified by PAR-CLIP to be targeted by miRNAs that were identified as being loaded more onto AGO2 in the combined RIP-seq/RNA-seq analysis..................139
Figure 32: Top GO-term analyses of genes identified by PAR-CLIP to be targeted by miRNAs that were upregulated in a p53-independent manner in the RNA-seq ..............................................................................143
Figure 33: Top GO-term and KEGG analyses of genes identified by PAR-CLIP to be targeted by miRNAs that were downregulated in a p53-dependent manner in the RNA-seq or that were identified as being loaded less onto AGO2 in the combined RIP-seq/RNA-seq analysis ..........144
Figure 34: Top GO-term analyses of genes identified by PAR-CLIP to be targeted by miRNAs that were downregulated in a p53-independent manner in the RNA-seq .................................................................145
Figure 35: CCND1 and POGZ are direct targets of let-7, miR-23a and miR-34a .................................................................146
Figure 36: A meme de novo motif discovery analysis .................................................................................................................................148
Figure 37: UCSC Genome Browser-adapted density plots ..........................................................................................................................149
Figure 38: DOX induces the expression of the GAS5-derived snoRNAs ........................................................................................................153
Figure 39: The GAS5-derived snoRNAs are not suitable housekeeping genes for use in experiments where DOX treatment ...................................................................................................................................................154
Figure 40: Expression of the GAS5-derived snoRNAs is greater in colorectal tumours than in benign colon tissue ..................................................................................................................................................................156
Figure 41: A comparison of p53 expression levels between paired normal colorectal and colorectal tumour tissue samples .....................................................................................................................................................157
Figure 42: A positive correlation exists between GAS5-derived snoRNA levels and p53 expression in colorectal tissue samples ............................................................................................................................................158
Figure 43: In microdissected colon samples the GAS5-derived snoRNAs are expressed more in malignant and pre-malignant tissue than benign tissue and levels correlate with p53 expression ...160
Figure 44: The GAS5-derived snoRNAs are not processed by DICER ..............................................................................................................................162
Figure 45: Bionanalser traces of cDNA derived from RNA produced in the PAR-CLIP experiment that was then used to create a cDNA library of small RNAs .................................................................167
Figure 46: Bionanalser traces of cDNA derived from RNA produced in the AGO2 IP experiment that was then used to create a cDNA library of small RNAs ..................................................................................168
Figure 47: Bionanalser traces of cDNA derived from RNA produced in the AGO2 IP experiment that was then used to create a cDNA library of mRNAs ........................................................................................................169
Figure 48: Complex network involved in p53-mediated regulation of CDKN1A .................................................................180
List of Tables

Table 1: miRNAs implicated in the pathogenesis of human malignancies........................................31
Table 2: Studies demonstrating the gene targets of the miR-34 family following transcriptional
induction by p53. ..........................................................................................................................49
Table 3: Studies demonstrating the gene targets of the miR-15a/16 following transcriptional induction
by p53 .........................................................................................................................................50
Table 4: Studies demonstrating the gene targets of miRNAs transcriptionally regulated by p53 ......52
Table 5: Determinants of miRNA-target interaction .......................................................................65
Table 6: List of mammalian cell lines..............................................................................................75
Table 7: Normal growth media ........................................................................................................75
Table 8: List of buffers and reagents ..............................................................................................76
Table 9: Primary antibodies ............................................................................................................78
Table 10: Secondary antibodies .....................................................................................................78
Table 11: List of luciferase 3’UTRs .................................................................................................79
Table 12: Transfection media and reagents ....................................................................................79
Table 13: Primers used for reverse transcription and their sequences ............................................79
Table 14: Summary of Samples. ....................................................................................................92
Table 15: A meme motif discovery analysis and TOMTOM analysis ............................................150
Abbreviations
-/-: p53 knockout cell lines
+/-: p53 wilt-type cell lines
4SU: 4-Thiouridine
5'TOP: 5'-terminal oligopyrimidine
6SG: 6-Thioguanosine
A: Adenine
A: AGO-IPed RNA
AAV: Adeno-associated virus
ACVR1B: Activin A Receptor Type IB
AGO2: Argonaute 2
AREG: Amphiregulin
ASO: Anti-sense oligonucleotide
Asp: Aspartate
ATM: Ataxia telangiectasia mutated
ATP: Adenosine triphosphate
ATR: Ataxia telangiectasia and Rad3 related
Bcl: B-cell lymphoma
BCSCs: Breast cancer stem cells
BMI1: polycomb ring finger oncogene
BMI1: Polycomb ring finger oncogene
C: Cytosine
Ca 19-9: Carbohydrate antigen 19-9
CBX3: Chromobox homologue 3
CCND1: Cyclin D1
CCNE2: Cyclin E2
CCR: Crosslink-centred region
cdc: Cell division cycle
CDC25B: Cell Division Cycle 25B
CDK4/6: Cyclin-dependent kinase 4/6
CDK6: Cyclin-dependent kinase 6
CDKN1A: Cyclin-dependent kinase inhibitor 1A
CDKN2A: Cyclin-dependent kinase inhibitor 2A
Chk: Checkpoint kinase 1
CK1: Casein kinase 1
CLIP: Crosslinking and immunoprecipitation
CLL: Chronic lymphocytic leukaemia
e-MYC: Myelocytomatosis viral oncogene homolog
Corr: Correlation
C-terminal: Carboxyl terminal
D: doxorubicin treated
DBD: DNA-binding domain
ddH2O: Double distilled water
DDX5: DEAD-box RNA helicase p68
DGCR8: Drosha and DiGeorge syndrome critical region gene 8
DKK1: Dickkopf Homolog 1
DLEU2: Deleted in lymphocytic leukaemia 2
DMEM: Dulbecco’s Modified Eagle’s Medium
DNA: Deoxyribonucleic acid
DOX: Doxorubicin
dsRBD: Double-stranded RNA-binding domain
dsRNA: Double stranded RNA
DTT: Dithiothreitol
E2F3: E2F transcription factor;
ECCs: Embryonal carcinoma cells
EDTA: Ethylenediaminetetraacetic acid
EIF5A: Eukaryotic Translation Initiation Factor 5A
EMT: Epithelial-mesenchymal transition
EREG: Epiregulin
ERα: Oestrogen receptor-alpha
ESPL1: Extra Spindle Poles-Like 1 Protein
EZH2: Enhancer of zeste homolog 2
FCS: Foetal calf serum
FGF: Fibroblast growth factor
FGF18: Fibroblast growth factor 18
FGFR: Fibroblast growth factor receptor
Fra-1: Fos-related antigen 1
FUS: Fused in Sarcoma
G: Guanine
G: IgG-Ipped RNA
Gadd45GIP1: Growth Arrest And DNA-Damage-Inducible Gamma Interacting Protein 1
GAS5: Growth arrest-specific transcript 5
GEP: Gene expression profiling
Glu: Glutamate
GTP: Guanosine triphosphate
h: Hour
HCC: Hepatocellular carcinoma
HCl: Hydrochloric
HCV: Hepatitis C virus
HDAC: Histone deacetylase
hESCs: Human embryonic stem cells
HIF-1 β: Hypoxia inducible factor β
HIF-1α: Hypoxia-inducible factor 1α
HIPK2: Homeodomain Interacting Protein Kinase 2
HIPK2: Homeo-domain interacting protein kinase-2
HITS-CLIP: High-throughput sequencing crosslinking immunoprecipitation
HMGA1: High mobility group AT-hook 1
HMGA2: High mobility group AT-hook 2
HNRNPC: Heterogenous Nuclear Ribonucleoprotein C
HSP: Heat shock protein
i.p: Intraperitoneal
i.v: Intravenous
I: input or total RNA
iCLIP: Individual-nucleotide resolution UV crosslinking and immunoprecipitation
IGF1: Insulin-like growth factor 1
IGF2BP1: Insulin-like growth factor 2 mRNA binding protein 1
IGFBP3: Insulin-like Growth Factor Binding Protein 3
IMPDH: Inosine 50-monophosphate dehydrogenase
IP: Immunoprecipitation
JNK1: Jun-NH2-terminal kinase-1
kb: Kilobases
KCl: Potassium chloride
kDA: Kilodaltons
KITLG: KIT ligand
KLF4: Kruppel-like factor 4
LAMC1: Laminin gamma 1
LDHA: Lactate dehydrogenase A
LEF1: Lef1 lymphoid enhancer binding factor 1
LEF1: Lef1 lymphoid enhancer binding factor 1
LEF1: Lymphoid Enhancer-Binding Factor 1
Let-7: Lethal-7
LNA: Locked nucleic acid
LRA: Luciferase reporter assay
LRP6: Low density lipoprotein receptor-related protein 6
LRP6: Low Density Lipoprotein Receptor-Related Protein 6 β-catenin
MAPK1: Mitogen-activated protein kinase 1
Mcl-1: Myeloid leukemia cell differentiation protein
MDM2: Murine double minute 2
MEFs: Mouse embryonic fibroblasts
MET: Hepatocyte growth factor receptor
MET: Mesenchymal-epithelial transition
Mid: Middle
miRISC: miRNA-containing RNA-induced silencing complex
miRNAs: microRNAs
mirtrons: Intron-derived miRNAs
MITF: Microphthalmia-associated transcription factor
mRNA: Messenger RNA
MRP-1: Multi-drug resistance-associated protein 1
MSPCs: Mammary stem/progenitor cells
MTA2: Metastasis associated 1 family, member 2
MTSS1: Metastasis suppressor 1
n: p53 knockout
NACC1: Nucleus Accumbens Associated 1
NaCl: Sodium chloride
NAD: Nicotinamide adenine dinucleotide
NaF: Sodium fluoride
ncRNAs: Non-coding RNAs
NOP56: Nucleolar protein 56
NSCLC: Non-small cell lung cancer
nt: Nucleotide
NUAK2: NUAK Family SNF1-Like Kinase 2
OD: Oligomerisation domain
OLFM4: Olfactomedin 4
ORF: Open reading frame
OS: Overall survival
PACT: Protein activator of PKR
PAR-CLIP: Photoactivatable-ribonucleoside-enhanced crosslinking & immunoprecipitation
PAZ: Piwi Argonaut and Zwille
PCNA: Proliferating Cell Nuclear Antigen
PCR: Polymerase chain reaction
PDAC: Pancreatic ductal adenocarcinoma
PFS: Progression-free survival
PIDD: p53-induced protein with death domain
PIP3: Phosphatidylinositol (3,4,5)-triphosphates
piRNAs: PIWI-interacting RNAs
PKMYT1: Protein Kinase Membrane Associated Tyrosine/Threonine 1
POGZ: Pogo transposable element with ZNF domain
Pol II: Polymerase II
POLL: Polymerase Lambda
PP2A: Pyrophosphatase 2
PRC1: Protein regulator of cytokinesis 1
pre-rRNA: Precursor rRNA
Pri-miRNA: Precursor miRNA
 pri-miRNAs: Primary microRNAs
PRR: Proline-rich region
PS: Prediction software
PTEN: Phosphatase and tensin homolog
PTP4A1: Protein Tyrosine Phosphatase Type IVA Member 1
PUMA: p53 upregulated modulator of apoptosis
Rb: retinoblastoma protein
RBM38: RNA-binding-motif protein 38
RBP: RNA binding protein
RIIDDs: RNase III domains
RIP: RNA-immunoprecipitation
RISC: RNA-induced silencing complex
RLC: RISC loading complex
RNA: Ribonucleic acid
RNAi: RNA inhibitor
RNase: Ribonuclease
RNP: Ribonucleoprotein
RRM2: Ribonucleotide reductase
rRNAs: Ribosomal RNAs
RT-qPCR: Reverse transcription quantitative polymerase chain reaction
scaRNAs: Small Cajal body-specific RNAs
SCC: Squamous cell carcinoma
sdRNAs: snoRNA-derived RNAs
SDS: Sodium Dodecyl Sulphate
SDS-PAGE: Sodium dodecyl sulphate polyacrilamide gel electrophoresis
Seq: Sequencing
Ser: Serine
SESN1: Sestrin 1
SIRT-1: Sitiuin-1
smarca1: SWI/SNF-Related Matrix-Associated Actin-Dependent Regulator Of Chromatin Subfamily A, Containing DEAD/H Box 1
SMC1A: Structural Maintenance Of Chromosomes 1A
SMIR: Small molecule inhibitor
snoRNA: Small nucleolar RNA
snRNA: Small nuclear RNA
T: Thymine
TAD: Transactivation domain
TBS: Tris-buffered saline
TBST: Tris-buffered saline tween
TGF: Transforming growth factor
Thr: Theonine
TRAIL: TNF-related apoptosis-inducing ligand
TRBP: TAR RNA-binding protein
TRIM28: Tripartite Motif Containing 28
TSP-1: Thrombospondin-1
TUBB: Type IB Class I Beta-Tubulin
TUSC2: Tumor Suppressor Candidate 2
U: Uridine
UBC: Ubiquitin C (UBC)
ULBP2: UL16 binding protein 2
UTR: Untranslated region
UV: Ultraviolet
V: vehicle treated.
VEGF: Vascular endothelial growth factor
VHL: von Hippel-Lindau
w: p53 wild type
WB: Western blot
WNT1: Wingless-type MMTV integration site family, member 1
WT: Wild-type
YY1: YY1 transcription factor
ZEB: Zinc-finger E-box binding homeobox
Chapter 1: Introduction
1.1 MicroRNAs and their biogenesis

MicroRNAs (miRNAs) are small endogenous non-coding RNAs ~17 – 25 nucleotides (nt) in length that play an important role in post-transcriptional regulation of gene expression by targeting mRNAs for cleavage or translational repression.\(^2\) The first miRNA gene, Lin-4, was identified in 1993 in the nematode Caenorhabditis elegans and was implicated in postembryonic development.\(^3\) \(^4\) miRNAs were subsequently identified in other species including plants, animals and even viruses.\(^5\) Early estimates grouped miRNAs into 48 families, based on similarities in nt 2–8 at their 5’ end, referred to as the ‘seed sequence’, and at least a third of miRNA families are highly conserved across species.\(^6\) \(^7\) Currently around 1900 mature human miRNA sequences have been identified and the genes that encode them have been found in intergenic regions and within introns and exons of other genes. miRNAs can be transcribed from their own promoters, promoters of nearby genes or from promoters of ‘host’ genes. Furthermore, they can be expressed as single genes (monocistronic) or as clusters of miRNAs situated within one specific locus (polycistronic).\(^4\)

MiRNA gene transcription is predominantly regulated by transcription factors (TF) responding to multiple stimuli.\(^8\) Most miRNAs are transcribed in the nucleus by RNA polymerase II (Pol II) to produce primary transcripts (pri-miRNAs).\(^9\) – \(^11\) Some miRNAs associated with Alu repeats are transcribed by Pol III, which is also responsible for transcription of other small non-coding RNAs such as the U6 small nuclear RNAs (snRNA).\(^2\) Pri-miRNAs are several kbs long, contain a local hairpin “stem-loop” structure and are capped and polyadenylated (Figure 1).\(^9\) – \(^11\)

Pri-miRNA processing involves “cropping” of the upper part of the pri-miRNA stem–loop by an RNaseIII enzyme called DROSHA, to form the 60-70-nt precursor-miRNA (pre-miRNA).\(^12\) \(^13\) Drosha is a 160kDa protein containing two RNase III domains (RIIIDs) and a double-stranded RNA-binding domain (dsRBD).\(^5\) \(^14\) It forms two types of complex, a small 650 kDa microprocessor complex that contains only Drosha and DiGeorge syndrome critical region gene 8 (DGCR8) that processes the majority of pri-miRNAs, and a larger complex that contains the Dead-Box RNA helicases (p72 and p68; also known as DDX17 and DDX5), heterogeneous nuclear ribonucleoproteins and Ewing's sarcoma proteins.\(^15\) p72 and p68 may act as a “scaffold” to recruit other factors (e.g. SMADs & p53) for the processing of a subset
of pri-miRNAs whose pre-miRNA and mature miRNA levels are reduced in homozygous p72 or p68 knockout mice.\textsuperscript{16-18}

Drosha cleaves the 5' and 3' arms of the pri-miRNA hairpin, and DGCR8 determines the exact cleavage site. Intron-derived miRNAs (mirtrons) bypass Drosha-mediated processing and form a hairpin structure resembling a pre-miRNA after splicing and debranching, prior to further processing by Dicer in the cytoplasm.\textsuperscript{19,20} Pre-miRNAs can also be generated from short introns, which are referred to as mirtrons. They are Drosha-independent and are liberated during splicing by the spliceosome.\textsuperscript{21}

Pre-miRNAs comprise of a short double-stranded RNA (dsRNA) stem with a base-paired 5' end and a 1–8 nt protruding 3’ end ("minihelix motif").\textsuperscript{9,22} Export of pre-miRNAs is mediated by a nuclear transport receptor called exportin-5, encoded by the XPO5 gene.\textsuperscript{23} Exportin-5 binds cooperatively to pre-miRNAs and the guanosine triphosphate (GTP) bound form of its cofactor Ran GTPase in the nucleus, and releases the pre-miRNA following the hydrolysis of Ran-GTP to Ran-GDP in the cytoplasm.\textsuperscript{23} Pre-miRNAs are then processed by the cytoplasmic RNase III Dicer which cleaves off the loop to form into 22 nt miRNA duplex (miRNA-miRNA*) that have 2 nts overhanging at each 3’ end.\textsuperscript{5,20}

Dicer is a 200 kDa multi-domain protein containing two RIIIDs, a dsRBD and a long N-terminal segment with a Dead-Box RNA Helicase domain, a DUF283 domain and a PAZ domain. The Dead-Box RNA Helicase domain uses ATP hydrolysis to unwind RNA duplexes.\textsuperscript{5} Human Dicer next interacts with 2 proteins, TRBP (TAR RNA-binding protein) and PACT (protein activator of PKR), neither of which are required for the cleavage reaction itself, but contribute to the formation of the RNA-induced silencing complex (RISC) by recruiting the final core component protein Argonaute-2 (AGO2).\textsuperscript{24,25} The RISC is the cytoplasmic effector machine of the miRNA pathway and contains a single-stranded miRNA guiding it to its target mRNAs. Cytoplasmic miRNA processing and RISC assembly are mediated by the RISC loading complex (RLC).\textsuperscript{20} The RLC complex is initiated by the assembly of Dicer, TRBP and AGO2, and the exported pre-miRNA only joins the RLC after this is complete.\textsuperscript{26}

Mature miRNAs that are incorporated into the effector complex are termed ‘miRNP’ (miRNA-containing ribonucleoprotein complex), “mirgonaut” or “miRISC” (miRNA-
containing RNA-induced silencing complex). The function requires separation of the miRNA/miRNA* duplex into the functional guide strand, which is complementary to the mRNA target, and the passenger strand (miRNA*), which is subsequently degraded. The strand selected for incorporation into the miRISC is determined by the thermodynamic stability of the base pairs at the two ends of the duplex. Generally the strand with the weakest 5'-end base pairing is selected as the functional strand and loaded onto AGO2. The miRISC then binds by imperfect or perfect base pair complementarity to the 3'UTR of the target mRNA in a sequence-specific manner, inducing mRNA cleavage (in cases of perfect base pair complementarity), cleavage-independent mRNA degradation through deadenylation or translational repression (in cases where the interaction contains bulges). MiRNAs generally bind via their 5' part, termed the ‘seed region’ (nucleotides 2–8 of the miRNA), to the complementary motif located on the target mRNA using Watson–Crick consecutive base pairing. The 3' part of the miRNA also interacts with the target thus forming an an imperfect interaction with it. It has been predicted that a single miRNA can therefore target many mRNAs. The choice between translational suppression or mRNA degradation is largely affected by the degree of complementarity with the mRNA target.

Although in plants, miRNAs generally target transcripts through perfect complementarity, in humans, targeting is predominantly through imperfect base-pairing to multiple sites in 3' UTRs. Watson–Crick base-pairing to the 5' end of miRNAs, especially to the ‘seed region’ is crucial for miRNA-target regulation. However, 3' end pairing might also contribute to target recognition, especially when sites have weaker miRNA seed matches. Imperfect miRNA–mRNA hybrids with central bulges at nucleotides 9–12, drive repression of gene expression through translational inhibition or exonucleolytic mRNA decay. Several studies in various systems have proposed mechanisms of miRNA-mediated translational repression. They include repression at the initiation and post-initiation stage of translation, as well as rapid degradation of target mRNAs. GW or P-bodies in the cytoplasm appear to play a vital role in miRNA-mediated translational repression and mRNA decay, and GW182 is a particularly important component of this process in various organisms including humans. Wakiyama et al. (2009) demonstrated that let-7 miRNPs directed the deadenylation of target mRNAs and impaired the synergistic enhancement of translation by the 5'-cap and 3'-poly(A) tail and that this resulted in translational repression and was dependent on AGO and GW182. Furthermore, Meijer et al. (2013) demonstrated a linear model for miRNA-mediated gene regulation in which translational repression via the RNA helicase eIF4A2 was
required initially and this was subsequently followed by mRNA decay. Perfectly complementary target sites with central pairing are rarely demonstrated in humans but generally result in target regulation through mRNA degradation/slicing rather than transcriptional repression. Target site multiplicity has been shown to enhance the degree of repression by animal miRNAs, each of which might have hundreds of target transcripts, owing to their imperfect base-pairing requirements. Although miRNAs predominantly target evolutionarily conserved sites on the 3’UTRs of mRNA, studies have shown that target sites also exist on the open reading frames (ORFs) as well as on the 5’ UTRs and that miRNA-guided gene repression via these sites can be equally efficient. Accordingly, two recent studies in human cells have identified functional miRNA target sites in the ORFs of DNA methyltransferase 3b and of the cell-cycle inhibitor p16INK4A (CDKN2A). Conversely, recent evidence from genome wide studies argues against miRNAs having the capability to regulate targets through an interaction with their ORFs region. Although miRNA regulated gene targeting relies predominantly on the miRNA-mRNA interaction sites within the primary structures of both elements, it is probable that secondary structure of the mRNA and its association with RNA-binding proteins (RBPs) could also influence target site accessibility, and hence modulate miRNA regulation of gene expression. For example, folding within the mRNA secondary structure (for example forming a stem loop) may conceal its 3’UTR target site and prevent the RISC from binding. Furthermore, if the miRNA target site overlaps with an RBP binding site then, when bound to the RBP, the mRNA would be protected from miRNA targeting.
1.1.1 AGO2 and miRNA loading on the RISC complex

AGO proteins are the direct binding partners of small RNAs and are crucial components of all of the small-RNA-guided gene-regulatory processes that have been identified to date.\textsuperscript{49} They are highly conserved in all eukaryotes, except in \textit{Saccharomyces cerevisiae}, where the small RNA machinery is lost.\textsuperscript{50} In eukaryotes, the AGO family can be divided into AGO and PIWI proteins. AGO proteins mainly interact with miRNAs or short interfering RNAs in the cytoplasm where they are involved in post-transcriptional gene-silencing.\textsuperscript{51} PIWI proteins silence transposable genetic elements through binding to PIWI-interacting RNAs (piRNAs) especially in germ line cells.\textsuperscript{52} In mammals, the AGO subfamily consists of four distinct proteins (termed AGO1-4), and although all four appear to bind miRNAs, only AGO2 is catalytically active and functions as an endonuclease.\textsuperscript{53} However studies do suggest that the
non-catalytic AGO proteins (AGO1, 3 and 4) may not be completely redundant and that they may play a role in processes such as spermatogenesis\textsuperscript{54} and \textit{Alu} RNA-guided mRNA decay during stem cell proliferation.\textsuperscript{55}

AGO2 consists of four distinct domains: the amino- or N-terminal, PAZ (PIWI–ARGONAUTE–ZWILLE), Mid (middle) and PIWI domains, that form a two lobed structure with one lobe comprising the N and PAZ domains and the other containing the MID and PIWI domains. The two lobes are connected via a hinge that allows structural rearrangements to occur upon RNA binding. The N domain facilitates small RNA loading onto AGO2 and assists in unwinding the small RNA duplex.\textsuperscript{56} Furthermore it contains an unstructured loop that is important for mRNA cleavage.\textsuperscript{57} The PAZ domain binds the miRNA 3’ end into a specific binding pocket and thus anchors it in place, whilst the MID domain anchors the 5’ end through specific contacts with uridine or adenine residues.\textsuperscript{58} This means that small RNAs with Gs or Cs at their 5’ end will only bind weakly to human AGO2.\textsuperscript{59} The PIWI domain contains an RNase-H-like fold, consisting of an Asp-Asp-Glu/Asp motif that is crucial for the endonuclease action of AGO2 and is responsible for target mRNA cleavage following fully complementary binding to the RISC.\textsuperscript{58}

The process of miRNA loading onto AGO in the RISC complex starts at the Dicer stage. Dicer binds pre-miRNAs and, via its two RNase III domains, cleaves both strands to form a dsRNA approximately 21 nt in length, which subsequently dissociates from and re-associates with Dicer at a different position. The dsRNA-binding protein, TRBP, is a crucial component in this step as it senses asymmetry and positions the dsRNA in a specific orientation that allows correct strand selection and loading. Therefore, Dicer contains two distinct RNA-binding sites: one that positions long dsRNA for cleavage, and one that rebinds the siRNA after cleavage, aided by TRBP.\textsuperscript{49} The dsRNA is then transferred to a bound AGO2 with the assistance of a chaperone protein called heat shock protein 90 (HSP90). HSP90 hydrolyses ATP and with the aid of other co-chaperones keeps AGO2 in an open state which allows it to accommodate the dsRNA. Following this, the passenger or miRNA* strand is removed, and the AGO protein loaded with the single-stranded miRNA transitions into a closed conformation.\textsuperscript{49} In humans, the miRNA* strand cleaved by AGO2 is removed from the RISC-loading complex by the endonuclease C3PO, which also activates the RISC, priming it
to be able to cleave complementary target mRNAs. Once bound, the miRNA then guides the RISC complex to its mRNA target.

### 1.1.2 The let-7 family of miRNAs

Lethal-7 (let-7) was the second miRNA ever to be discovered (Lin-4 being the first) and was originally demonstrated as a heterochronic gene in the nematode Caenorhabditis elegans, where it was shown to control the timing of stem-cell division and differentiation.\(^{60}\) Let-7 subsequently became the first known human miRNA,\(^{61}\) and studies have shown that let-7 and its family members are highly conserved across species in both sequence and function.\(^{7,62}\) Furthermore, in humans, aberrant expression of let-7 creates a less differentiated cellular state and promotes the development of a number of diseases such as cancer.\(^{7}\)

The let-7 family of miRNAs is often present in multiple copies in a genome although this varies between organisms. For example, although there is only one let-7 in Drosophila, there are 11 mature let-7 sequences in the Zebrafish and 14 let-7 family members in mice.\(^{7}\) To distinguish between multiple isoforms, a letter is placed after let-7 to indicate a slightly different nucleotide sequence, and a number at the end denotes that the same sequence is present in multiple genomic locations. In humans, there are 10 mature let-7 family sequences (let-7a to g, let-7i, mir-98, and mir-202) that are produced from 13 closely related genes/precursor sequences (let-7a-1, let-7a-2, let-7a-3, let-7b, let-7c, let-7d, let-7e, let-7f-1, let-7f-2, let-7g, let-7i, mir-98, and mir-202).\(^{7}\) Therefore, 3 separate precursors (let-7a-1, let-7a-2 and let-7a-3) produce the mature let-7a sequence and precursors from two different genomic locations produce the let-7f (let-7f-1 and let-7f-2) sequence.

Let-7 is an important regulator of seam cell differentiation and development in C. elegans and in particular the transition from the larval 4 to adult stage, which is the point at which these cells stop dividing and terminally differentiate.\(^{60,63}\) Let-7 mutant worms display abnormalities in temporal development patterns and many die, thus giving the let-7 gene its name, lethal-7.\(^{60}\) Interestingly, let-7 expression levels vary at the different stages of C.elegans development and this regulation appears to occur at the transcriptional level. This is crucial for regulating the proper timing of seam cell development especially at the last larval stage; let-7 is only detectable from the L3 stage of development and reaches maximal expression.
during the L4 stage,\textsuperscript{60} consistent with its role in promoting cell-cycle exit and terminal differentiation of seam cells at the end of the L4 stage. This temporal pattern of let-7 expression during development has also been reported in other organisms,\textsuperscript{64-66} although the direct contribution of let-7 in development in vertebrates has not yet been demonstrated.\textsuperscript{65,67} This relates to the fact that in vertebrates some let-7 family members are probably redundant, and the fact that it is technically challenging to knock out all let-7 family members in the same animal at the same time.\textsuperscript{7} However, let-7 has been shown to play a role in embryogenesis and cerebral development in mammals\textsuperscript{65,67} and expression levels are undetectable in human and mouse embryonic stem cells but increase with cellular differentiation.\textsuperscript{7,65} Conversely, low let-7 levels have been linked to a number of human malignancies which is consistent with the process of cellular de-differentiation that is commonly associated with tumorigenesis.\textsuperscript{68-70} However, some let-7 family members appear to have an oncogenic role in certain tumour types which suggests that the function of this family of miRNAs may be sequence and cell-type specific. For example, in lymphoma, upregulation of let-7b and let-7i was associated with high grade transformation and hence a worse outcome.\textsuperscript{71}

Although let-7 expression is regulated at the transcriptional level in C.elegans, in humans it has been shown to be under the control of both transcriptional and post-transcriptional mechanisms. Here, let-7 genes are generally transcribed by Pol II leading to the formation of pri-let-7. The transcription of let-7a, let-7d, let-7f and let-7g, is directly repressed by MYC,\textsuperscript{72} although in humans, few transcription factors have been identified that promote let-7 transcription.\textsuperscript{7} In contrast a number of transcription factors, including temporal regulatory element-binding protein and DAF-12 are thought to regulate the transcription of let-7 in C.elegans.\textsuperscript{73,74} However, human let-7 is also post-transcriptionally regulated by the RNA-binding protein, LIN28, which binds to the stem loops of pri-let-7a and pri-let-7g,\textsuperscript{75,76} and inhibits the Drosha complex from processing them.\textsuperscript{76,77} In some cases let-7a processing is inhibited by LIN28 at the Dicer level rather than the Drosha level, through the direct binding of LIN28 to the pre-let-7a stem, which prevents the production of mature let-7a.\textsuperscript{78} Interestingly, as cells differentiate, the expression of LIN28 decreases and this enables the Drosha complex to bind to and process pri-let-7 into pre-let-7, which is then processed by Dicer into the mature let-7.\textsuperscript{77} This is then incorporated into the RISC where it can post-transcriptionally repress its gene targets. A number of let-7 targets have been identified and include the cell-cycle regulators cyclin D,\textsuperscript{79} TLX,\textsuperscript{79} CDC25A,\textsuperscript{80} RAS,\textsuperscript{81} MYC\textsuperscript{82} and the high
mobility group AT-hook 2 (HMGA2), and it is regulation of these genes that enables cells to terminally differentiate. In targeting such genes let-7 appears to have an important tumour suppressor role and this is supported by the fact that decreased expression of many let-7 family members has been associated with a number of malignancies and a poor prognosis in many cases. Therefore improving our knowledge of how let-7 expression is regulated and identifying, in greater detail, the network of let-7 regulated genes may enable a better understanding of its role in tumorigenesis.

1.2 miRNAs and cancer

Over 50% of miRNA-encoding loci reside in chromosomal regions altered by tumorigenesis and a number of miRNAs function as classical oncogenes or tumor suppressor genes. The first study to demonstrate a link between miRNAs and cancer identified both miR-15a and miR-16-1 were down-regulated or absent in most patients with B-cell chronic lymphocytic leukemia, which resulted from a deletion at 13q14 where the genes encoding these miRNAs are located. Extensive tumor profiling studies have implicated many miRNAs in the development, progression and metastasis of many tumour types (Table 1). Their discovery goes someway to explaining the gap that frequently exists between tumour genotype and phenotype, and has furthered our understanding of post-transcriptional regulation of gene expression in which they play a critical role. The functional importance of miRNAs in both healthy and pathological states has exposed their notable potential as disease biomarkers and therapeutic targets.
<table>
<thead>
<tr>
<th>Tumour type</th>
<th>miRNA</th>
<th>Expression in pathological state</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Breast cancer</td>
<td>miR-155</td>
<td>Overexpressed</td>
<td>87</td>
</tr>
<tr>
<td></td>
<td>miR-21</td>
<td>Overexpressed</td>
<td>88</td>
</tr>
<tr>
<td></td>
<td>miR-17-92</td>
<td>Overexpressed</td>
<td>89</td>
</tr>
<tr>
<td></td>
<td>miR-9</td>
<td>Overexpressed</td>
<td>90</td>
</tr>
<tr>
<td></td>
<td>miR-31</td>
<td>Down-regulated</td>
<td>91</td>
</tr>
<tr>
<td></td>
<td>let-7</td>
<td>Down-regulated</td>
<td>92</td>
</tr>
<tr>
<td></td>
<td>miR-126</td>
<td>Down-regulated</td>
<td>93</td>
</tr>
<tr>
<td>Colorectal cancer</td>
<td>miR-155</td>
<td>Overexpressed</td>
<td>94</td>
</tr>
<tr>
<td></td>
<td>miR-21</td>
<td>Overexpressed</td>
<td>95</td>
</tr>
<tr>
<td></td>
<td>miR-34</td>
<td>Down-regulated</td>
<td>96</td>
</tr>
<tr>
<td>Lung cancer</td>
<td>miR-21</td>
<td>Overexpressed</td>
<td>97</td>
</tr>
<tr>
<td></td>
<td>miR-17-92</td>
<td>Overexpressed</td>
<td>98</td>
</tr>
<tr>
<td></td>
<td>miR-155</td>
<td>Overexpressed</td>
<td>99</td>
</tr>
<tr>
<td></td>
<td>let-7</td>
<td>Down-regulated</td>
<td>99</td>
</tr>
<tr>
<td></td>
<td>miR-1</td>
<td>Down-regulated</td>
<td>100</td>
</tr>
<tr>
<td>Hepatocellular carcinoma</td>
<td>miR-221</td>
<td>Overexpressed</td>
<td>101</td>
</tr>
<tr>
<td></td>
<td>miR-21</td>
<td>Overexpressed</td>
<td>102</td>
</tr>
<tr>
<td></td>
<td>miR-151</td>
<td>Overexpressed</td>
<td>103</td>
</tr>
<tr>
<td></td>
<td>miR-26a</td>
<td>Down-regulated</td>
<td>104</td>
</tr>
<tr>
<td>Prostate cancer</td>
<td>miR-21</td>
<td>Overexpressed</td>
<td>105</td>
</tr>
<tr>
<td></td>
<td>miR-34a</td>
<td>Down-regulated</td>
<td>106</td>
</tr>
<tr>
<td></td>
<td>miR-15a/16-1</td>
<td>Down-regulated</td>
<td>107</td>
</tr>
<tr>
<td></td>
<td>miR-143</td>
<td>Down-regulated</td>
<td>108</td>
</tr>
</tbody>
</table>
### Table 1: miRNAs implicated in the pathogenesis of human malignancies

<table>
<thead>
<tr>
<th>Cancer</th>
<th>miRNAs</th>
<th>Expression</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pancreatic cancer</td>
<td>miR-21, miR-34, let-7</td>
<td>Overexpressed, Down-regulated</td>
<td>109, 110, 22</td>
</tr>
<tr>
<td>AML</td>
<td>miR-29a, miR-10a/b, miR-29b</td>
<td>Overexpressed, Down-regulated</td>
<td>111, 112, 113</td>
</tr>
<tr>
<td>CLL</td>
<td>miR-155, miR-15a/16-1, miR-34a</td>
<td>Overexpressed, Down-regulated</td>
<td>114, 86, 115</td>
</tr>
</tbody>
</table>

### 1.2.1 MiRNAs as biomarkers

#### 1.2.1.1 Diagnostic biomarkers

Cancers are often diagnosed at a late stage, with associated poor prognosis. The oncogenic and tumor suppressive nature of miRNAs, and the discovery of tumour-specific miRNA signatures suggests a potentially important role for these molecules as early diagnostic biomarkers. In a study of 104 matched pairs of primary malignant and non-malignant lung tissue, Yanaihara *et al.*<sup>99</sup> identified a group of 43 differentially expressed miRNAs that could successfully discriminate between the two groups. Gee *et al.*<sup>116</sup> found a panel of miRNAs (including the miR-200 family) that were down-regulated in malignant pleural mesothelioma compared to lung adenocarcinoma, and could be used consistently to distinguish between these two tumors.<sup>116</sup> A tissue profiling study demonstrated a role for miRNAs in distinguishing primary brain tumours from secondary metastases originating from distant sites<sup>117</sup> and other studies have identified panels of potential diagnostic biomarkers in hepatocellular carcinoma,<sup>118</sup> breast carcinoma<sup>90</sup> and pancreatic endocrine and acinar tumours.<sup>119</sup>
Current diagnostic methods are usually invasive and technically challenging, indicating a substantial need for novel non-invasive biomarkers for early tumor detection. The ability to profile miRNAs in the circulation represents a less invasive method of investigating disease-specific miRNAs and is a promising alternative approach to tumor tissue profiling techniques. An essential requirement for developing circulating miRNA-based diagnostics is the ability to accurately isolate and measure miRNA species. Chen et al.\textsuperscript{120} was one of the first studies to demonstrate the presence of miRNAs in human serum and plasma. Using Solexa sequencing they identified approximately 100 circulating miRNAs in healthy Chinese subjects and subsequently studied specific expression patterns of serum miRNAs in lung and colorectal cancer patients, comparing them to healthy subjects. In lung cancer patients, 28 miRNAs were absent and 63 new miRNA species were detected, and in the colorectal cohort 69 serum miRNAs were detected that were not present in the healthy cohort, which included miR-221, previously shown to be increased in colorectal tumour specimens.\textsuperscript{121} Interestingly, despite the high concentration of RNAses in plasma and serum, circulating miRNAs were resistant to RNase A digestion.

Evidence for the role of secretory miRNAs as diagnostic tools is growing. Elevated plasma miR-155, miR-197, and miR-182 levels accurately discriminated lung cancer patients from healthy controls (81.33% sensitivity and 86.76% specificity). miR-155 and miR-197 levels were higher in patients with metastasis than those without, and were significantly decreased in responsive patients during chemotherapy.\textsuperscript{122} Shen et al.\textsuperscript{123} demonstrated that patients with malignant solitary pulmonary nodules had elevated plasma levels of miR-21 and miR-210 but lower miR-486-5p levels compared to those with benign lesions or healthy controls. A comparison of plasma miRNA expression levels between 20 early breast cancer patients and 20 healthy controls identified 31 differentially expressed miRNAs in Caucasian patients and 18 in African American patients.\textsuperscript{124} Liu et al.\textsuperscript{125} used logistic modeling to show plasma miR-16 and miR-196a levels could discriminate pancreatic cancer from chronic pancreatitis and normal controls. This was even more sensitive with the inclusion of serum CA19-9 in the logistic model, and was even effective at identifying stage I disease.

Despite this growing evidence, the use of plasma miRNAs as diagnostic tools in clinical practice remains sparse. Although there have been advancements in isolation techniques and despite the apparent stability of circulating miRNAs, their measurement may be confounded by variability in the levels of cellular miRNAs from hematological origins and from
circulating tumor cells. To truly gain a greater understanding of the relevance of free circulating miRNA levels, fractionation processes must be employed to differentiate between those derived from tumor tissue and circulating cells. This should improve the sensitivity of these methods, as microdissection has in tissue-based miRNA biomarker studies. Furthermore, miRNA levels in other bodily fluids may play a future role in cancer diagnosis. Studies have identified miRNAs as potential diagnostic biomarkers in human bile\textsuperscript{126}, sputum \textsuperscript{127} and faeces\textsuperscript{128} although further studies are required to validate this.

1.2.1.2 Prognostic biomarkers

miRNA profiling studies have demonstrated a number of correlations between differentially expressed miRNAs and prognosis in various tumour types. In lung adenocarcinoma, low let-7a and elevated miR-155 expression was associated with poor survival\textsuperscript{99}, and miR-137 and miR-372 levels correlated with increased risk of relapse and worse survival\textsuperscript{129}. Overexpression of miR-155 and miR-21 correlated with poor outcome in early-breast cancer and dysregulated expression of miR-9, miR-10b, miR-21 are miR-315 was associated with an increased risk of metastasis\textsuperscript{90,91,93,130}. Schetter \textit{et al.}\textsuperscript{131,132} demonstrated in two independent cohorts that elevated miR-21 levels correlated with poor survival in colorectal cancer, and the same miRNA has been implicated in predicting survival in localised pancreatic cancer, as has miR-196a-2\textsuperscript{109}. In hepatocellular carcinoma decreased miR-122\textsuperscript{133} and miR-26\textsuperscript{104} correlated with poor patient survival and miR-151-5p with an increased risk of intrahepatic metastasis\textsuperscript{134}. Historically, factors such as tumor grade, size and lymph node involvement have been used in early stage disease to determine treatment strategies; however there is sufficient evidence to suggest that miRNAs would serve as useful adjuncts or even alternatives to such methods.

1.2.1.3 Predictive biomarkers

Despite the development of novel targeted anti-cancer therapies, chemotherapy represents the foundation for treatment regimens for most hematological and solid malignancies. However, resistance of tumor cells to chemotherapy, and to a lesser extent targeted therapies, remains a major obstacle to effective treatment. Recent studies postulate that aberrant miRNA expression might be involved in tumor resistance to current therapies. This suggests a role for
miRNAs as predictive biomarkers, and that modulation of tumor miRNAs may be exploited to improve treatment response in addition to producing direct anti-tumor effects.

miR-155 knockdown increased sensitivity of breast cancer cells to chemotherapy through regulation of FOXO3a and down-regulation of miR-21 augmented breast cancer cell response to taxol. Multidrug-resistance in specific MCF-7 cell lines is associated with reduced levels of miR-326 and miR-451 leading to upregulated expression of the multidrug resistance-associated protein 1 (MRP-1/ABCC1) and the multi-drug resistance 1 protein (MDR1) respectively. Furthermore, reconstitution of miR-451 expression sensitised cells to doxorubicin. Inhibition of miR-21 increased sensitivity of pancreatic adenocarcinoma cells to gemcitabine. Two clinical studies provide further evidence that miR-21 affects chemosensitivity in pancreatic cancer. The first study using 81 pancreatic ductal adenocarcinoma (PDAC) samples from patients treated with gemcitabine, found high miR-21 expression was associated with poorer overall survival in both the adjuvant and metastatic settings, and a subsequent study demonstrated a correlation between low miR-21 expression and improved outcome (disease-free and overall survival) in patients with localised PDAC treated with adjuvant gemcitabine or 5-fluorouracil chemotherapy.

Cisplatin resistance has been linked to miR-214 overexpression via targeting of PTEN. Li et al. demonstrated an association between taxol-resistance in human ovarian cancer cells lines and increased expression of MDR1/P-glycoprotein due to down-regulation of miR-27a, and transfection with pre-miR27a re-sensitized these cells to taxol. Furthermore, in a study of 37 stage III ovarian cancer patients, seven miRNAs, including miR-27a, were significantly differentially expressed in tumors from platinum-resistant versus -sensitive patients. High miR-27a expression was associated with a particularly poor prognosis in terms of OS.

Response to fludarabine therapy in CLL is associated with differential miRNA expression. In a study of fludarabine-treated CLL patients, Ferracin et al. identified 37 miRNAs that distinguished responders from non-responders, with miR-21, miR-148a and miR-122 being more highly expressed in non-responding patients. In a similar study involving 50 CLL patients, fludarabine resistance was associated with decreased miR-29a and increased miR-181a expression. The Let-7 family targets Bcl-X̂ in hepatocellular carcinoma cell lines, and overexpression of these miRNAs increased sorafenib-induced apoptosis in cell culture experiments. Furthermore, in a miRNA expression analysis involving over 200 HCC
patients treated with interferon-α, Ji et al.\textsuperscript{104} demonstrated an association between improved response and low miR-26 expression.

Although these data are promising, larger prospective trials are required to validate the role of miRNAs as predictive biomarkers, but such studies may lead to significant changes in treatment algorithms for certain tumour types.

1.2.2 MiRNA-modulating agents as cancer therapeutics

1.2.2.1 Theory versus reality

A greater understanding of miRNA expression and function, and the growing evidence that miRNA deregulation is involved in cancer development and progression supports their role as potential therapeutic targets in cancer. Down-regulation of target oncogenes by re-expression of tumor suppressor miRNAs, or re-expression of tumor suppressor genes by silencing an oncomir could impair tumour growth and metastasis.

A single miRNA can target many mRNAs suggesting that miRNA-modulating therapy could simultaneously modify a number of relevant gene networks within a tumour, leading to significant biological effects on phenotype and possibly a reduction in the emergence of therapy-resistant clones. Furthermore, we have recently shown for the first time that in situations where multiple miRNAs co-operatively inhibit a network of tumour suppressor genes to promote tumorigenesis, co-targeting more than one miRNA is more effective in suppressing tumorigenesis tumour growth than targeting just a single miRNA.\textsuperscript{147}

Accordingly, a new class of drugs that specifically target small RNA pathways via replacement of tumor suppressive miRNAs with synthetic or viral vector encoded miRNA mimics or antisense-mediated inhibition of oncogenic miRNAs are currently in development. Hypothetically, such approaches should be well tolerated by non-target cells, because pathways within them that are affected by miRNA-modulating therapy are already controlled by the endogenous miRNA. However, a number of concerns must be addressed before such therapies can be safely applied to the clinic. Off-target effects provide a major obstacle, although utilizing a number of miRNAs at lower concentrations to target a single mRNA may enhance the specificity of silencing, and selecting a miRNA which targets multiple genes within a pathway may consolidate silencing and reduce off-target effects.\textsuperscript{148} Furthermore,
although double-stranded RNA-based therapies have been shown to be more effective than single stranded mimetics, they trigger a greater innate immune response through activation of double-stranded RNA-dependent protein kinase R.\textsuperscript{149} This effect can be reduced by the careful design of double-stranded based therapies less than 30 nucleotides in size that are relatively deplete in uracil or guanine sequences.

The development of approaches to deliver miRNA-modulating agents to target tissues is also a major difficulty. Barriers to systemic delivery include degradation by serum and tissue nucleases, failure to cross the capillary endothelium due to size, uptake by scavenger macrophages and ineffective endocytosis and endosomal release in target cells.\textsuperscript{150} Additionally, intracellular RNA-binding proteins may further limit the activity of miRNA-modulating agents within target cells. The strategies employed to modulate miRNA activity for therapeutic purposes and measures utilised to overcome potential obstacles are discussed below.

1.2.2 miRNA inhibition

miRNA antagonists must selectively hybridize with their endogenous miRNA target via partial or complete complementarity, thereby preventing interactions between the miRNA and its target mRNA.\textsuperscript{151} The most basic examples are anti-miRNA oligonucleotides (AMOs) which consist of a ‘naked’ single-stranded molecule that inhibits miRNAs via complementary binding. An early example of their use \textit{in vivo} demonstrated that intravenous (i.v.) injection of an AMO into mice silenced hepatocyte expression of Fas and protected against fulminant hepatitis.\textsuperscript{152}

However ‘naked’ oligomers are relatively unstable and are easily degraded by endogenous RNases, resulting in their limited efficacy by systemic administration. To overcome this, oligomers can be modified by the addition of cholesterol conjugated 2’-O-methyl groups to produce more stable ‘antagomirs’. Krutzfelt \textit{et al.}\textsuperscript{153} demonstrated that a single i.v. injection of an antagomir designed to target cholesterol-regulating miR-122 in mice, resulted in prolonged miR-122 silencing in the liver and a significant decrease in serum cholesterol levels. Ma \textit{et al.}\textsuperscript{154} showed the systemic delivery of an antagomir to miR-10b in a mouse mammary tumour model prevented metastasis formation. Furthermore, a single intratumoral injection of antagomir-221/222 into Me665/1 melanoma xenografts in nude mice,
significantly inhibited tumour progression for one week with no documented toxicity, and intraperitoneal (i.p.) injection of antagomir-182 reduced hepatic metastasis of melanoma cells in a mouse model.

Further adaptations led to the development of ‘locked nucleic acid’ (LNA) oligomers. Such oligonucleotides contain a ribose moiety that is functionally locked into a C3’-endo conformation via the addition of a methylene bridge, that confers greater stability, increased miRNA-binding affinity and lower toxicity. LNA antisense oligomers to miR-122 have been shown to reduce serum cholesterol levels in healthy and obese mice as well as healthy non-human primates. MiR-122 is also essential for Hepatitis C virus (HCV) RNA replication and systemic delivery of an LNA antisense oligomer to miR-122 (SPC3649, Santaris Pharma) in HCV-infected chimpanzees led to prolonged 300-fold suppression in HCV viremia. In a subsequent Phase I single-dose safety study in humans, SPC3649 demonstrated limited toxicity and a clear dose-dependent pharmacology and has now entered into Phase II, making it the first miRNA-modulating therapy to reach this stage. Such agents are yet to show similar success in the cancer setting.

Vector-encoded RNA molecules, termed ‘miRNA sponges’, represent a novel approach to miRNA-modulating therapy. Containing multiple complimentary 3’UTR binding sites, they competitively bind to miRNAs thus liberating their mRNA targets. They can be designed to carry a number of different binding sites, enabling simultaneous inhibition of multiple members of a miRNA cluster or different miRNAs acting on the same target. This is an advantage over ASOs which only target single miRNAs. Furthermore they can be stably integrated into the genome, enabling the development of transgenic animals and stable cell lines that are functionally deficient in certain miRNAs. Valastyan et al. orthoptically implanted MCF7-Ras cells expressing a sponge vector targeting the anti-metastatic miR-31 into mice, resulting in a significant induction of lung metastases. Gentner et al. demonstrated that expression of an anti-miR-223 vector in hematopoietic stem cells, resulted in the functional knockdown of miR-223 when these cells were transplanted into lethally irradiated mice. However, a number of factors make miRNA sponges unsuitable for therapeutic use in humans and therefore no such approaches have been trialed yet. Firstly there is the risk of insertional mutagenesis in target cells, and vector size and poor biodistribution limits their systemic use. The future design of small-molecule drugs targeting miRNAs (SMIRs) may overcome this.
1.2.2.3 miRNA replacement/mimics

Although studies directed at inhibiting oncogenic miRNAs have shown promise, the restoration of tumour suppressive miRNAs using miRNA replacement or mimics may be a more efficacious, less toxic strategy. They are likely to have less off-target effects, as miRNA mimics should have a similar sequence to the naturally occurring depleted miRNA and therefore target identical mRNAs. Developing a miRNA mimic requires the design of structures able to interact with the RISC and target the same mRNA as the endogenous miRNA. Such approaches have utilised chemically modified miRNAs, viral vectors and nanoparticle delivery systems in pre-clinical, and more recently, clinical models.

Chemical adaptations to miRNAs mimics have, as with miRNA antagonists, allowed more effective administration to their target tissue. Takeshita et al.\textsuperscript{160} used tail vein injections to administer a chemically modified miR-16 precursor (a tumor suppressor known to be down-regulated in prostate cancer) or ‘scrambled mimic’ to a murine model of bony metastatic prostate cancer. The miRNA was complexed with atelocollagen, which promoted uptake into the bone metastases and was effectively and persistently detected in target tissue for >3 days. Growth of the bone metastases was significantly lower in miR-16-treated mice than those administered the ‘scrambled mimic’.

Adeno-associated viruses (AAVs) allow the persistent transcription and expression of miRNAs at high levels in target tissues with a low risk of insertional mutagenesis compared to other viral delivery systems.\textsuperscript{161} Down-regulation of miR-26a is associated with HCC and tail vein delivery of a miR-26a-expressing AAV into a murine HCC model suppressed tumorigenesis.\textsuperscript{162} Furthermore, although approximately 90% of hepatocytes were transduced with miR-26a in this model, there were no signs of hepatotoxicity or dysregulation of endogenously expressed miRNAs.\textsuperscript{162} Another murine model utilised the intranasal instillation of an adenovirus encoding let-7 or a negative control (n.c.) miRNA, and the cre recombinase in transgenic K-RAS G12D mice (which induces expression of the K-RAS mutant G12D and the formation lung tumors). Following surgical removal of the lungs, histology revealed mice that received cre/let-7 developed far fewer and smaller tumors than those that received cre/n.c., further establishing a role for let-7 as a tumor suppressor.\textsuperscript{163} A previous study had shown that the intratumoral injection of let-7 directly into murine non-small cell lung cancer
xenografts caused tumor shrinkage, demonstrating two methods of replacing this downregulated miRNA. AAVs may carry the risk of undesired immune responses and to date, no such therapeutic strategies have been employed in human cancer trials.

Nanoparticles are positively charged structures with diameters of 45-70nm that can be used to administer negatively charged miRNAs/mimics to target tissues. This structure confers greater miRNA stability, allows their slow release for prolonged mRNA targeting and avoids the possible immunogenicity associated with AAVs. These characteristics suggested their potential to be administered intravenously to humans. Davis et al. recently published the preliminary results of the first human clinical trial using this delivery system. This Phase I study of patients with advanced solid tumors used the iv administration of an RNAi encapsulated in nanoparticles that targets the mRNA of ribonucleotide reductase (RRM2), a protein overexpressed in many solid tumors. The nanoparticles also contained surface transferrin protein targeting ligands (present on tumor cell surfaces) allowing specific delivery to tumors. The study demonstrated effective uptake of the RNAi to target tissue and efficient knockdown of its target gene. Response data are not yet mature.

The clinical relevance of miRNAs is clearly reflected by the fact that in the ten years since initially being linked to malignancy, they have progressed from discovery to biomarker and drug development programs and remain at the forefront of research into tumour biology. Their role as biomarkers is promising particularly with regards to circulating miRNAs, which offer a potentially less invasive method of diagnosing cancer, assessing risk of relapse, and predicting and potentially following response to therapy. Guidance regarding the design of biomarker studies is rightly becoming more stringent and additional appropriately planned studies involving larger sample sizes will be vital before specific miRNAs can be utilised clinically in this manner. miRNA-modulating agents represent a new class of therapeutics, encompassing a wide range of mechanistic approaches including RNA interference and gene therapy, as well as complex delivery and tissue-targeting strategies. It is these last two points that have proved the biggest obstacle in the development of such agents. The human body’s natural barriers have hampered the systemic delivery of these drugs although this is being slowly overcome by innovative modifications to current agents and the design of new therapeutic structures. The generally widespread expression of many pertinent miRNAs has required the design of novel delivery systems to ensure tissue-specific targeting. Initial results from work by Davis et al. and developments in delivery systems by the pharmaceutical
industry suggest that before long, the use of miRNA-based therapies may become common practise in the cancer clinic. Furthermore, linking evidence from prognostic and predictive biomarker studies with clinical trials involving miRNA-modulating therapies could lead to combined treatment strategies involving the use of such therapies with current treatments to maximise response and improve outcome.

1.3 The importance of regulatory relationships between p53 and miRNAs in the p53-mediated cellular response to DNA damage

p53 is one of the most frequently mutated tumour suppressors, and functions predominantly as a transcription factor by regulating target gene expression and to a lesser extent through non-transcriptional processes such as those mediated by protein-protein interactions or by its direct binding with the mitochondria.\(^{166,167}\) It regulates the expression of protein-coding genes and non-coding RNAs involved in apoptosis, senescence, DNA repair, cell-cycle arrest, cell differentiation, metabolism and angiogenesis,\(^ {168}\) thereby coordinating cellular response to stresses, such as oncogene activation and DNA damage. Whether the cell enters growth arrest or undergoes apoptosis ultimately depends on the integration of incoming signals with antagonistic effects on cell growth.\(^ {168}\) Mutations or inactivation in p53 are the most frequent abnormalities observed in cancer cells.\(^ {169}\) Furthermore, aberrant expression of proteins involved in the p53 pathway such as murine double minute 2 (MDM2),\(^ {170}\) Jun-NH\(_2\)-terminal kinase-1 (JNK1),\(^ {171}\) hypoxia-inducible factor 1\(\alpha\) (HIF-1\(\alpha\))\(^ {172}\) and imbalances in Rb-E2F1 and Bax/BCL-2 ratios are commonly observed in malignant cells.\(^ {168}\) Such mutations have significant implications in terms of tumor growth and response to current treatment modalities. Many chemotherapeutic agents, in particular anthracyclines and platinum-containing compounds, as well as radiotherapy, induce cancer cell death or senescence via DNA damage and the subsequent activation of p53-mediated pathways. Resistance to such treatments can occur due to mutations in components of the p53-signalling networks. A better understanding of the role of the various components within these pathways and their interactions with each other may allow modifications and improvements to current anti-cancer treatments and the design of novel therapies. Refining our knowledge of the role of miRNAs in such p53-signalling networks may be crucial to achieving this.
1.3.1 p53 activation and its regulation of downstream signaling pathways following DNA damage.

In humans, the TP53 gene is located on the short arm of chromosome 17 (17p13.1) and spans 20 kb, with a non-coding exon 1 and a very long first intron of 10 kb. The coding regions are highly conserved between vertebrates. TP53 encodes a protein that in humans comprises of 393 amino acids and is separated into 7 domains consisting of folded DNA-binding and tetramerization domains, flanked by intrinsically disordered regions at both the amino- and carboxy-termini each of which has a specific function. The first 63 residues form the natively unfolded amino-terminal transactivation domain (TAD), which is subdivided into the subdomains TAD1 (residues 1-42) and TAD2 (residues 43-61), which are primarily responsible for activating transcription factors and connect target gene recognition with target gene expression by direct binding to the transcriptional coactivators p300/CBP and components of the basal transcription machinery. The TAD region is followed by a proline-rich region (PRR; residues 61-94) which is important in apoptosis. Next comes the central DNA-binding core domain (DBD; residues 94-292) that consists of a central immunoglobulin-like β-sandwich scaffold and additional elements that form the DNA-binding surface, which include a loop-sheet-helix motif and two large loops stabilized by a zinc ion. Residues 316-325 comprise the nuclear localization signaling domain and residues 307-355 contain the homo-oligomerisation domain (OD) that is connected to the DBD through a flexible linker region. The C-terminal is located between residues 356-393 and contains sequences involved in nuclear localisation, tetramerisation and the regulation of specific and non-specific DNA binding. This regulatory region, along with the amino-terminal, undergo a variety of post-translational modifications including ubiquitylation, phosphorylation, methylation and SUMOylation that regulate activity and function of the protein.

p53 operates primarily as a transcription factor with both transactivation and repressor functions and is biologically active as a homotetramer that cooperatively binds to its target duplex DNA in a sequence-specific manner. The target binding sites were originally believed to only consist of two decameric half-site motifs formed of RRRCWGYYY (R= adenine or guanine; W= adenine or thymine; Y= cytosine or thymine) separated by 0–13 base pairs and usually located 5’ to the gene or in the first or second intron of the gene regulated by p53. However, recent studies have also shown that many p53 response elements have
consecutive half-sites and that an increase in spacer length between half-sites for a given response element correlates with a decrease in p53 affinity and transactivation.\textsuperscript{187,188} Furthermore, Wang et al. (2009) developed a set of predictive rules to distinguish between response elements accountable for activation or repression of target genes, the latter having characteristic deviations from the canonical response element sequence that are generally associated with weaker affinity.\textsuperscript{189}

In its role as a transcriptional regulator, p53 controls the expression of hundreds of genes involved in mediating cell growth, division and survival, apoptosis and DNA repair, angiogenesis, metabolic processes and even the immune response.\textsuperscript{190,191} Its activity is often transient and this is regulated mainly through changes in its protein stability.\textsuperscript{181} p53 is induced and activated by a range of stress stimuli including DNA damage, hypoxia, heat/cold shock, activated oncogenes (hyper-proliferation), nutritional deprivation, ribosomal stress, nitric oxide and spindle damage.\textsuperscript{17,18} Each stress stimulus induces p53 by suppressing its degradation and this leads to a coordinated cellular response that occurs in a context-dependent manner, the outcome of which is reliant upon the balance between growth arrest, senescence and DNA repair and that of apoptosis or programmed cell death, and appears specific to the cell type and the manner, intensity and duration of the activating stress.\textsuperscript{17,18} To add further complexity to the process, p53 can also operate in a transcription-independent manner and regulate the onset of apoptosis through a direct interaction with the mitochondrion and its apoptotic machinery.\textsuperscript{167} The purpose of such a complex but coordinated p53-mediated cellular response to stress is to protect cells from damage induced by the stimulus and to allow them to repair or reverse the insult where possible or ‘hibernate’ until the stimulus diminishes, but also to allow the cell to undergo programmed cell death in the event of irreparable or prolonged damage. Perhaps one of the most important roles of p53 is to coordinate tumour suppression mechanisms that block uncontrolled proliferation of transformed or damaged cells.\textsuperscript{17,18}

Levine et al. (2006)\textsuperscript{17} elegantly described the p53 pathway by dividing it conveniently into five parts:

1. The input signals that induce the network (e.g. DNA damage)
2. The upstream regulators that detect and interpret those signals and activate the functional pathway by relaying the inputs to the p53 protein or molecules that most
immediately regulate its concentration and activity (e.g. ataxia telangiectasia mutated; ATM).

3. The core set of proteins, including the p53 itself, which regulates p53 activity and function (e.g. MDM2).

4. The downstream events which are composed of sets of genes and their proteins that are regulated by p53, most commonly by transcriptional activation but in some cases by protein–protein interactions (e.g. CDKN1A)

5. The cellular outputs of these downstream events (e.g. cell cycle arrest, cellular senescence or apoptosis).

DNA damage is probably the best described ‘input signal’ in this context and most relevant to this thesis. DNA damage takes the form of single- or double-strand breaks and initiates the p53 response predominantly through the activation of the upstream regulators ataxia telangiectasia and Rad3 related (ATR) and ATM, respectively. These are protein kinases that coordinate and assimilate various signalling pathways leading to changes in the post-translational status of p53 itself and several of its direct or indirect regulators such as MDM2 and MDMX.\(^{192-194}\) MDM2 is the principal regulator of p53 protein levels and function. Through its role as a RING-finger type E3 ligase, MDM2 can mono- or poly-ubiquitylate lysine residues on p53 leading to nuclear export or proteosomal degradation of p53, respectively.\(^{195,196}\) MDMX, a functionally defective E3 ligase related to MDM2 also binds the MDM2/p53 complex and promotes MDM2-mediated ubiquitylation and suppresses p53-mediated transcription.\(^{197,198}\) Therefore, ATR/ATM-mediated post-translational phosphorylation of p53 leads to its almost universal uncoupling from MDM2-induced degradation and allows the subsequent recruitment of crucial transcription factors, leading to chromatin remodelling and transcriptional activation.\(^{17}\)

For example, following a double strand break, ATM (and later ATR) phosphorylate serine residues on the N-terminus of p53 (such as Serine 15 or Serine 20), and these phosphorylation events initiate a cascade of further post-translational modifications. The protein kinase, casein kinase 1 (CK1), uses the phosphorylation of Ser15 as a recognition determinant and sequentially phosphorylates Threonine (Thr) 18 which then inhibits p53/MDM2 association and promotes the uncoupling of p53 from degradation and nuclear export.\(^{199,200}\) Ser20 phosphorylation occurs through the activation of the CHK2 protein kinase by ATM, which further uncouples p53 from MDM2 mediated suppression. Phosphorylation
events do not only affect p53 signalling through the promotion of its uncoupling from regulatory partners such as MDM2, but they also do so through the regulation of transcription factor/co-activator recruitment. For example, Thr18 phosphorylation promotes the binding of the p300/CBP translational co-activators which stimulate p53 transcriptional activity. Furthermore, phosphorylation of Ser46 by various protein kinases, including ATM, homeo-domain interacting protein kinase-2 (HIPK2) and protein kinase C-delta following DNA damage, has been suggested to be involved in promoting apoptosis through induction in the expression of genes such as PTEN. ATM and ATR also activate a number of other proteins following DNA damage that induce other post-translational modifications on p53 and its regulators. These include deubiquitinases such as HAUSP which deubiquitylates and activates p53, phosphatases such as pyrophosphatase 2 (PP2A) that remove phosphate groups from MDM2 and p53 thus suppressing p53 activity, neddylases such as NEDD8 which inhibit p53 transcriptional activity, and deacetylases such as histone deacetylase (HDAC) and situin-1 (SIRT-1) that provide crucial p53 deacetylation activity to shut off p53-dependent transcription. Interestingly, DNA damage has been show to induce the acetylation of p53 at various lysine residues on the C-terminus by acetyltransferases, and this promoted p53 stability and favoured survival over cell death by promoting p53-mediated activation of cell cycle arrest genes.

The downstream events mediated by p53 following a stress stimulus usually involves a large number of genes with the precise gene signature being specific to the nature and intensity of the stimulus and the cell type and surrounding environment. Some p53-regulated genes appear to be transcribed in response to many different types of stress signals and in all tissues responding to the stress (e.g. CDKN1A, cyclin G, Growth Arrest and DNA damage 45 (GADD45) and MDM-2) and others appear to be neither stimulus- nor tissue specific (e.g. PTEN). Following DNA damage, p53 can mediate cell cycle arrest by inducing a number of negative regulators of the cell cycle such as CDKN1A, 14-3-3 sigma and GADD-45, and can activate apoptosis through the upregulation of positive mediators of both the intrinsic and extrinsic apoptotic pathways. These include Fas, TRAIL, killer/DR5, p53-induced protein with death domain (PIDD), bax, noxa and puma. p53 can also induce senescence through the transcriptional upregulation of CNKN1A, which activates the retinoblastoma protein (Rb) to shut down the transcription of E2F target genes. DNA repair can also be mediated by p53 through its regulation of the expression of proteins such as p48, the p53R2 subunit of ribonucleotide reductase, and the sestrins.
Interestingly, as well as regulating response pathways in a cell that has undergone DNA damage, p53 can also regulate the ability of that cell to communicate with other cells in its environment through the transcription of secreted proteins. In this way a cell can ‘inform’ neighbouring cells that it has DNA damage and coordinate an appropriate response in its surrounding environment. For example, p53-mediated upregulation and secretion of Insulin-like Growth Factor Binding Protein 3 (IGFBP3) prevents the activation of growth signalling pathways in neighbouring cells. Similarly the p53-regulated protein thrombospondin is secreted following DNA damage and can alter the cellular matrix and suppress angiogenesis.\textsuperscript{228,229} Finally, as well as regulating the signaling pathways described above, many p53-responsive genes also initiate positive or negative feedback loops with the p53 protein and the core gene products of the pathway.\textsuperscript{230} For example, in some cell types, PTEN, which is induced by p53, dephosphorylates Phosphatidylinositol (3,4,5)-triphosphate (PIP3) leading to a reduction in the activity of the AKT kinase. AKT kinase normally phosphorylates MDM2, resulting in the translocation of MDM2 into the nucleus where it inactivates p53, but this is inhibited by PTEN here. This is an example of a positive feedback loop whereby p53 activity is enhanced through the actions of the genes that it transcriptionally induces.\textsuperscript{230,231} Other genes involved in similar pathways include CDKN1A,\textsuperscript{230} cyclin G,\textsuperscript{232} WIP-1\textsuperscript{233,234} and MDM2.\textsuperscript{235,236}

1.3.2 The regulatory relationships between p53 and miRNAs

miRNAs have evolved to allow cells to effectively respond to stresses such as DNA damage, hypoxia and oncogene activation.\textsuperscript{237} Accordingly, they have been strongly associated with the p53 stress-response pathway, and in recent years, have been identified as key mediators of p53-associated tumor suppression. Interactions between p53 and miRNAs occur at various stages. At the transcriptional level, p53 can induce the expression of several miRNAs which mediate its role in apoptosis, epithelial-mesenchymal transition (EMT), stemness, angiogenesis, senescence and cell cycle arrest.\textsuperscript{238-240} Interestingly, p53 can also directly repress expression of oncogenic miRNAs, such as the miR-17-92 cluster.\textsuperscript{241} These p53-regulated miRNAs have the potential to be involved in complex feedback and feedforward loops and contribute collectively to control, amplify and fine-tune cellular signals in response to stress, although there a few examples of this in the literature. Furthermore, gene targets of p53-regulated miRNAs may also be transcriptionally regulated.
by p53, creating even more complex coherent and incoherent feedforward loops. p53 is able to regulate miRNA expression transcriptionally, by a direct interaction with a subset of pri-miRNA promoters, and also post-transcriptionally. Post-transcriptionally, it regulates miRNA processing and maturation by interacting with the Drosha processing complex, associating with the DEAD-box RNA helicase p68. miRNAs can also negatively regulate p53 protein levels through direct binding to the p53 mRNA 3’-UTR, and can positively regulate p53 function through targeting several negative regulators of p53. miRNAs could therefore potentially act as novel p53 target genes to facilitate p53-induced tumor suppression. Furthermore, they may be as important in regulating p53 activity as the previously identified phosphatases (e.g. Wip1), kinases (e.g. ATR, ATM) and ubiquitin ligases (Cop1, MDM2). Additionally, the effects of epigenetic and genetic variations in the p53-miRNA interactome further suggest the importance of these networks in tumorigenesis.

### 1.3.2.1 Transcriptional Regulation of miRNAs by p53

p53 is a transcription factor and mainly exerts its regulatory action by a direct interaction with protein-coding target gene promoters. However, studies over the last 5 years have shown that non-coding RNAs, such as miRNAs, are also under its direct transcriptional control. Furthermore, there appears to be variability in the p53-regulated miRNA expression signatures, and their gene targets, depending on the cell type, nature of the stress, and cause of DNA damage.

In 2007, several groups using various screening methods reported miR-34a and miR34b/c as the first identified miRNA genes to be p53-regulated. These and subsequent studies demonstrated that miR-34a/b/c display multiple tumor suppressive activities and are summarized in Table 2. Bioinformatically predicted and experimentally validated miR-34 targets comprise cell cycle-related factors, predominantly those involved in G1/S phase transition, including E2F3, cyclin E2, cyclin-dependent kinase 4/6 (CDK4/6) and MYC. Furthermore, miR-34 has been shown to target genes involved in EMT (e.g. SNAIL), DNA repair, angiogenesis, stemness (e.g. CD44, WNT1/3), cell survival (e.g. SIRT1), metabolism (e.g. LDHA) and tumor immune surveillance. Accordingly, decreased expression of miR-34 has been observed in various tumors, suggesting that its loss
could promote tumorigenesis and further implying its potential importance in stabilizing p53-signalling.

Other miRNAs shown to be transcriptionally induced by p53 include miR-15a/16, miR-107, miR-122, miR-143, miR-145, miR-182, the miR-192/194/215 and miR-200 family members, miR-205, miR-605 and miR-1204. Although all these miRNAs have been shown, in at least one study, to be induced by direct binding of p53 to the relevant miR-gene promoter, it is worth noting that not all of the studies demonstrating potential or validated gene targets of these miRNAs specifically show this in each case. Interestingly, miR-22 has been shown to be a direct transcriptional target of p53 following doxorubicin-induced DNA damage, and the resulting upregulation in the expression of this miRNA led to targeting of CDKN1A and stimulated cells to undergo apoptosis rather than cell cycle arrest.

Genes involved in p53-signalling pathways that are targeted by miR-15a/16 are shown in Table 3. These miRNAs are encoded by an intron of the DLEU2 gene and although they were initially thought to be only induced by p53 post-transcriptionally, Fabbri et al. (2011) subsequently demonstrated the direct transcriptional regulation of DLEU2 by p53. The majority of studies demonstrating that miR-15a/16 target genes involved in p53-signalling do not specifically demonstrate direct binding of p53 to the miRNA promoter. However, these miRNAs have been implicated in targeting genes involved in various p53-associated signaling pathways including G0/G1-S transition, cell proliferation, migration, invasion and clonogenicity in various cancer cell lines. miR-15a/16-1 deletions have been detected in several tumor types including, CLL, non-small cell lung cancer and prostate cancer, suggesting that an impaired p53 response may drive them.

The miR-200 family has been strongly associated with p53-regulated signaling pathways that also play a crucial role in suppressing the formation of metastases, including EMT, stemness, clonogenicity and angiogenesis, through direct targeting of genes including ZEB1/2, SOX2, Klf4 and VEGF. miR-107 contributes to the role of p53 in the regulation of angiogenesis and hypoxic signaling through the targeting of hypoxia inducible factor-1β (HIF-1β), which interacts with HIF-1α subunits to form a transcription factor complex that mediates the tumor response to hypoxia. miR-107 is also involved in the regulation of G1/S cell cycle arrest through its gene target CDK6. Its expression is inversely associated with
HIF-1β levels in human colon cancer, and ectopic expression of miR-107 suppresses both tumor growth and angiogenesis in murine colon cancer models.\textsuperscript{259,269}

p53 binds directly to a p53 response element in the miR-145 promoter to induce expression of this tumor suppressive miRNA. miR-145 targets and negatively regulates c-MYC, and other cell cycle regulators such as CDK4/6, which accounts in part for the miR-145-mediated inhibition of tumor cell proliferation \textit{in vitro} and \textit{in vivo}.\textsuperscript{240,270} miR-145 represses pluripotency in human embryonic stem cells (hESCs) through the targeting of OCT4, SOX2, and KLF4,\textsuperscript{271} and regulates MDM2 in squamous cell carcinoma (SCC) HN30 and SCC-25 cells leading to increased p53-induced apoptosis and senescence.\textsuperscript{272} The tumor suppressive role of other p53-induced miRNAs is summarized in Table 4.

Interestingly some miRNAs are transcriptionally down-regulated, rather than induced, by p53 although relatively few have been identified or experimentally validated to date. These include miR-100, miR-125b, miR-99b, miR-221 and miR-501.\textsuperscript{253} Let-7a/b expression was down-regulated in a p53-dependent manner in colon cancer HCT116 cells exposed to DNA damage through UV radiation.\textsuperscript{273} Let-7a/b levels were inversely correlated with the expression of pro-apoptotic BAX and PUMA, although direct targeting of these genes was not shown in this study. Exogenous expression of let-7a/b increased radiation-induced cytotoxicity in a p53-dependent manner. These results were the first demonstration of a mechanistic connection between the radiation-induced stress response and the regulation of miRNAs and radiation-induced cytotoxicity.\textsuperscript{273} Furthermore, although most p53-regulated miRNAs function as tumor suppressors, some act as oncogenes. miR-194 is transcriptionally induced by p53 in colon cancer HCT116 cells and targets thrombospondin-1, leading to increased tumor angiogenesis.\textsuperscript{274} Furthermore, in melanoma cells, p53 directly up-regulates miR-149*, that in turn targets glycogen synthase kinase-3α, resulting in increased expression of Mcl-1 and resistance to apoptosis. This provides further evidence that miRNA*s (passenger strands) play important roles in regulating gene expression similar to their complementary mature miRNAs.\textsuperscript{275}
<table>
<thead>
<tr>
<th>miRNA</th>
<th>Up or down reg by p53</th>
<th>Related Gene targets</th>
<th>Gene target validation method</th>
<th>Cellular pathway</th>
<th>Cell/organism</th>
<th>Ref</th>
</tr>
</thead>
<tbody>
<tr>
<td>miR-34a/b/c</td>
<td>Up</td>
<td>CDK4/6, CCNE2, MET</td>
<td>WB corr, LRA</td>
<td>Cell cycle arrest</td>
<td>MEFs</td>
<td>259</td>
</tr>
<tr>
<td>miR-34a</td>
<td>Up</td>
<td>E2F3, NOTCH1, DLL1</td>
<td>PS</td>
<td>Apoptosis, G1 arrest</td>
<td>H1299, MCF7</td>
<td>253</td>
</tr>
<tr>
<td>miR-34a</td>
<td>Up</td>
<td>IMPDH</td>
<td>WB/RTqPCR corr</td>
<td>Nucleotide biosyn, GTP-dependent Ras signalling</td>
<td>H1299, HCT116</td>
<td>276</td>
</tr>
<tr>
<td>miR-34a/b/c</td>
<td>Up</td>
<td>SNAIL, SLUG, ZEB1/2, BM11, CD44, CD133, OLFM4, c-MYC</td>
<td>WB corr, LRA</td>
<td>Stemness, TGFβ signalling, MET/EMT,</td>
<td>H1299, HCT116</td>
<td>254</td>
</tr>
<tr>
<td>miR-34a/b/c</td>
<td>Up</td>
<td>SNAIL, β-catenin, LEF1, Axin2</td>
<td>WB Corr, LRA</td>
<td>Invasion, EMT/EMT</td>
<td>HCT116, MCF7, A549</td>
<td>277</td>
</tr>
<tr>
<td>miR-34a/b/c</td>
<td>Up</td>
<td>β-catenin, WNT1, WNT3, LRP6, LEF1</td>
<td>LRA, PS</td>
<td>WNT signalling, stemness, invasion</td>
<td>A549, MCF7, Xenopus</td>
<td>278</td>
</tr>
<tr>
<td>miR-34a</td>
<td>Up</td>
<td>FRA-1</td>
<td>LRA, PS, WB corr</td>
<td>Migration, invasion</td>
<td>HEK293, RKO, HCT116</td>
<td>279</td>
</tr>
<tr>
<td>miR-34a</td>
<td>Up</td>
<td>c-MYC, c-MET, c-Myc–Skp2–Miz1, c-Myc–pTEFB</td>
<td>LRA, PS</td>
<td>c-myc activated transcription, invasion</td>
<td>PC-3</td>
<td>280</td>
</tr>
<tr>
<td>miR-34b/c</td>
<td>Up</td>
<td>Myc, CDK6, Delta-like1, NOTCH1, Met, Ezh2, E2F3, Bcl2, Cyclin D1</td>
<td>PS</td>
<td>Proliferation, adhesion-independent growth</td>
<td>Primary mouse OSE cells</td>
<td>251</td>
</tr>
<tr>
<td>miR-34a/c</td>
<td>Not shown</td>
<td>ULBP2</td>
<td>LRA, PS, WB corr</td>
<td>Tumor immune surveillance</td>
<td>Natural Killer Cells</td>
<td>256</td>
</tr>
<tr>
<td>miR-34a</td>
<td>Up</td>
<td>AXL, LEF1, MTA2, LDHA, YY1</td>
<td>LRA, genome wide pSILAC, microarray</td>
<td>Glycolysis, lipid metabolism, WNT-sign</td>
<td>SW480</td>
<td>255</td>
</tr>
<tr>
<td>miR-34a</td>
<td>Up</td>
<td>SIRT1</td>
<td>LRA, corr WB</td>
<td>Apoptosis</td>
<td>HCT116</td>
<td>247</td>
</tr>
</tbody>
</table>

Table 2: Studies demonstrating the gene targets of the miR-34 family following transcriptional induction by p53.
CCNE2: cyclin E2; CDK4/6: cyclin dependent kinase 4/6; MET: hepatocyte growth factor receptor; MEFs: mouse embryonic fibroblasts; corr: Correlation; WB: western blot; LRA: luciferase reporter assay; PS: Prediction software; RTqPCR: realtime quantitative PCR; IMPDH: Inosine 50-monophosphate dehydrogenase; EMT: epithelial-mesenchymal transition; MET: mesenchymal-epithelial transition; Fra-1: Fos-related antigen 1; ZEB1/2: zinc finger E-box binding homeobox 1/2; BM11: polycomb ring finger oncogene; OLFM4: olfactomedin 4; c-MYC: myelocytomatosis viral oncogene homolog; LEF1: Lef1 lymphoid enhancer binding factor 1; WNT1/3: wingless-related MMTV integration site 1/3; ULBP2: UL16 binding protein 2; SIRT1: sirtuin 1; MTA2: metastasis associated 1 family, member 2; LDHA: lactate dehydrogenase A; YY1: YY1 transcription factor; E2F3: E2F transcription factor; BCL2: B-cell CLL/lymphoma 2; EZH2: enhancer of zeste homolog 2 (Drosophila)
<table>
<thead>
<tr>
<th>miRNA</th>
<th>Up or down reg(^a) by p53</th>
<th>Related Gene targets</th>
<th>Gene target validation method</th>
<th>Cellular pathway</th>
<th>Cell/organism</th>
<th>Ref</th>
</tr>
</thead>
<tbody>
<tr>
<td>miR-15a/16</td>
<td>Not shown</td>
<td>FGF2, FGFR1, WNT3A, CCND1, BCL-2</td>
<td>PS, LRA</td>
<td>Proliferation, migration, invasion</td>
<td>CAFs, PC3, CAHPV10, DU145</td>
<td>268</td>
</tr>
<tr>
<td>miR-15a/16</td>
<td>Up by p53/E2F1</td>
<td>CCNE</td>
<td>Corr WB</td>
<td>G1/S transition, cell cycle progression</td>
<td>H1299, U2OS</td>
<td>264</td>
</tr>
<tr>
<td>miR-15a/16</td>
<td>Not shown</td>
<td>BMI1</td>
<td>PS, corr WB, LRA</td>
<td>Cell proliferation, clonogenicity</td>
<td>Ovarian cancer, OVCAR-5, OV-202</td>
<td>265,267</td>
</tr>
<tr>
<td>miR-15a/16</td>
<td>Not shown</td>
<td>CDK4/6, CCND1/3, CCNE1</td>
<td>PS, corr WB</td>
<td>G1/S transition</td>
<td>A549, H2009, H129, H358</td>
<td>266</td>
</tr>
<tr>
<td>miR-15a/16</td>
<td>Not shown</td>
<td>CCND2/3, Cdk4/6, Chk1</td>
<td>PS, GEP</td>
<td>G0/G1-S phase transition</td>
<td>B-cells/CLL</td>
<td>107</td>
</tr>
<tr>
<td>miR-15a/16-1</td>
<td>Not shown</td>
<td>CCND1, WNT3a</td>
<td>LRA, corr WB</td>
<td>Survival, proliferation, invasion</td>
<td>LNCaP, RWPE-1</td>
<td>107</td>
</tr>
</tbody>
</table>

Table 3: Studies demonstrating the gene targets of the miR-15a/16 following transcriptional induction by p53

CCNE2: cyclin E2; CDK4/6: cyclin dependent kinase 4/6; CCND1: cyclin D1; corr: Correlation; WB: western blot; LRA: luciferase reporter assay; PS: Prediction software; RTqPCR: real-time quantitative PCR; FGF: fibroblast growth factor; FGFR: fibroblast growth factor receptor WNT1/3; wingless-related MMTV integration site 1/3; BCL2: B-cell CLL/lymphoma 2; Chk: checkpoint kinase 1
<table>
<thead>
<tr>
<th>miRNA</th>
<th>Up or down reg&lt;sup&gt;d&lt;/sup&gt; by p53</th>
<th>Related Gene targets</th>
<th>Gene target validation method</th>
<th>Cellular pathway</th>
<th>Cell/organism</th>
<th>Ref</th>
</tr>
</thead>
<tbody>
<tr>
<td>Let7a/let7b</td>
<td>Down</td>
<td>Inverse correlation</td>
<td>Corr</td>
<td>p53-dependent response to DNA damage UV radiation</td>
<td>HCT116</td>
<td>273</td>
</tr>
<tr>
<td>miR-107</td>
<td>Up</td>
<td>CDK6, ?p130</td>
<td>LRA, PS</td>
<td>Regulation G1/S arrest</td>
<td>HCT116</td>
<td>209</td>
</tr>
<tr>
<td>miR-107</td>
<td>Up</td>
<td>HIF-1β</td>
<td>LRA, PS, corr WB</td>
<td>Angiogenesis</td>
<td>HCT116</td>
<td>259</td>
</tr>
<tr>
<td>miR-122</td>
<td>Up</td>
<td>Activates AKT</td>
<td>Corr WB</td>
<td>AKT/p53 apoptotic circuit</td>
<td>Cutaneous T-cell lymphoma</td>
<td>281</td>
</tr>
<tr>
<td>miR-143/145</td>
<td>Up</td>
<td>MDM2</td>
<td>LRA, corr WB/RTqPCR</td>
<td>p53 activation, Apoptosis /cell cycle arrest</td>
<td>HN30, SCC-25</td>
<td>272</td>
</tr>
<tr>
<td>miR-143/145</td>
<td>Up</td>
<td>K-Ras, MYC, CCND2, cdk6, E2F3</td>
<td>LRA, corr WB/RTqPCR</td>
<td>G1-growth arrest</td>
<td>HCT116</td>
<td>270</td>
</tr>
<tr>
<td>miR-145</td>
<td>Up</td>
<td>-</td>
<td>LRA, corr WB/RTqPCR</td>
<td>Apoptosis</td>
<td>Prostate Ca e.g. PC3</td>
<td>282</td>
</tr>
<tr>
<td>miR-145</td>
<td>Up</td>
<td>c-MYC</td>
<td>LRA, corr WB/RTqPCR</td>
<td>Growth inhibition</td>
<td>HCT116, MCF7</td>
<td>240</td>
</tr>
<tr>
<td>miR-145</td>
<td>Not shown</td>
<td>OCT4, SOX2, and KLF4</td>
<td>LRA, corr WB</td>
<td>Repress pluripotency in hESCs</td>
<td>hESCs</td>
<td>271</td>
</tr>
<tr>
<td>miR-149&lt;sup&gt;*&lt;/sup&gt; (oncogenic)</td>
<td>Up</td>
<td>GSK3α (Mcl-1 indirectly)</td>
<td>LRA, corr WB</td>
<td>Resistance to apoptosis (oncogenic)</td>
<td>Melanoma, Mel-RM</td>
<td>275</td>
</tr>
<tr>
<td>miR-182</td>
<td>Up</td>
<td>MITF, BCL2, CCND2</td>
<td>LRA, corr WB/RTqPCR</td>
<td>Cell growth, invasion, migration</td>
<td>Uveal melanoma M23, SP6.5</td>
<td>283</td>
</tr>
<tr>
<td>miR-182</td>
<td>Not shown</td>
<td>CCNE2, CDK2/4/6 MTSS1</td>
<td>LRA, corr WB</td>
<td>Metastasis</td>
<td>HCC, HLE, HLF, HepG2, Hep3B, HUH1</td>
<td>284</td>
</tr>
<tr>
<td>miR-192/215</td>
<td>Up</td>
<td>Multiple, siRNAs, LRA</td>
<td>LRA, PS, corr WB</td>
<td>Cell proliferation, G1/S &amp; G1/M transition</td>
<td>A549, TOV21G, HCT116</td>
<td>285</td>
</tr>
<tr>
<td>miR-192,194,215</td>
<td>Up</td>
<td>IGF1</td>
<td>LRA, PS, corr WB</td>
<td>Cell survival/migration</td>
<td>Myeloma cell lines</td>
<td>286</td>
</tr>
<tr>
<td>miR-192</td>
<td>Up</td>
<td>DHFR</td>
<td>LRA, corr WB/RTqPCR</td>
<td>Cell cycle control, cell proliferation</td>
<td>HCT116</td>
<td>277</td>
</tr>
<tr>
<td>miR-192</td>
<td>Up</td>
<td>ZEB2</td>
<td>LRA, corr WB/RTqPCR</td>
<td>EMT/migration/invasion</td>
<td>HCC cell lines</td>
<td>261</td>
</tr>
<tr>
<td>miR-194&lt;sup&gt;*&lt;/sup&gt; (oncogenic)</td>
<td>Up</td>
<td>TSP-1</td>
<td>LRA, corr WB/RTqPCR</td>
<td>Promotes angiogenesis</td>
<td>HCT116</td>
<td>278</td>
</tr>
<tr>
<td>miR-200 family</td>
<td>Up</td>
<td>ZEB1/ZEB2</td>
<td>LRA</td>
<td>EMT/migration/invasion</td>
<td>HCC, Breast cancer cell lines</td>
<td>261,262</td>
</tr>
<tr>
<td>miR-200c</td>
<td>Up</td>
<td>BMI1</td>
<td>PS, LRA</td>
<td>Inh stemness, clonogenicity &amp; tumor growth</td>
<td>Human BCSCs, Human/murine MSPCs, ECCs</td>
<td>288</td>
</tr>
</tbody>
</table>
Table 4: Studies demonstrating the gene targets of miRNAs transcriptionally regulated by p53

| miR-200c | Up | ZEB1, Sox2 and Klf4 | PS, LRA | Inh stemness & tumor growth | PANC1, MiaPaCa2, A549 |
| miR-200b | Not shown | VEGF | LRA | Angiogenesis | Lung CAFs, lung cancer cell lines |
| miR-200 | Not shown | VEGF, VEFG1 | LRA, PS, WB corr | Growth, invasion, metastasis |
| miR-205 | Up | E2F1, LAMC1 | PS, LRA, corr WB/RTqPCR | cell proliferation, cell cycle progression, clonogenicity, cell adhesion/migration | MDA-MB-231, BT-549, MCF7 |
| miR-205 | Not shown | ZEB1/ZEB2 | LRA | EMT/migration | Breast cancer cell lines, MDCK cells |
| miR-605 | Up | MDM2 | LRA, corr WB | Apoptosis | A549, MCF-7, HCT116 |
| miR-1204 | Up | - | - | Apoptosis, proliferation, cell cycle arrest | HCT116, RKO, SK-HEP1 |

1.3.2.2 Post-transcriptional regulation of miRNA expression and function by p53

p53 controls miRNA expression post-transcriptionally through the regulation of miRNA processing and maturation. This was first demonstrated in a study by Suzuki et al. (2009), where the authors used quantitative polymerase chain reaction with reverse transcription (RT-qPCR) to measure miRNA expression levels in p53<sup>+/−</sup> HCT116 cells following DNA-damage with doxorubicin. They observed an increase in miR-15a, miR-16-1, miR-23a, miR-26a, miR-103, miR-143, miR-203 and miR-206, in addition to the previously published miR-34a. However, although DNA damage induced the expression of the pri-miRNA, pre-miRNA and mature miR-34a, several other miRNAs (including miR-15a, miR-16-1, miR-143 and miR-145) increased only at the precursor and mature levels. Thus suggesting a role of p53 in enhancing the post-transcriptional maturation of these miRNAs, which all have
important roles in the regulation of cell cycle transition and cell proliferation. The maturation of these miRNAs was induced by the association of p53 with DROSHA and its partner DGCR8 via the interaction with DEAD-box RNA helicase p68 (DDX5) that is part of this complex (microprocessor). This enhanced DDX5 interaction with the microprocessor and promoted the cleavage of pri-miRNAs to pre-miRNAs. Furthermore, a transcriptionally inactive mutant p53 impeded the binding of p68 to the complex leading to a reduction in miRNA processing which may have relevance to tumor biology. Interestingly, DNA mutations in the p53 gene that are repeatedly observed in tumors (e.g. R273H and R175H) are associated with decreased DROSHA-mediated processing of pri-miRNAs and reduced expression of their mature miRNAs. Furthermore other members of the p53 family, such as p63 and p73, appear to regulate components of miRNA processing (i.e. Drosha-DGCR8, Dicer-TRBP2, and Argonaute proteins). A recent study also found a link between DNA damage-induced activation of p53 and miRNA biogenesis. Here, the investigators demonstrated that following DNA damage, phosphorylation of Serine-46 on p53 led to transcriptional upregulation of amphiregulin (AREG) through binding to its promoter region. AREG was shown to interact with DDX5 and regulate processing of pre-miRs such as pre-miR-15a which led to a reduction in the expression of the anti-apoptotic protein, Bel-2.

In addition to promoting miRNA maturation through DROSHA processing, the p53 response also monitors miRNA maturation at the level of DICER1. Loss of DICER1 function can be tumorigenic, predominantly due to incomplete miRNA maturation. DICER1 deficiency and subsequent loss of mature miRNAs in embryonic fibroblasts, induced expression of p53 and ARF, leading to reduced proliferation and premature senescence, which could be rescued by the deletion of p53 or Cdkn2a, which encodes ARF. This provides evidence that efficient miRNA processing is monitored by a p53-dependent checkpoint, and that loss of p53 function may significantly increase the tumorigenic potential of cancer cells with deficient DICER1 function or reduced mature miRNA levels.

Finally, p53 can also influence miRNA target gene selection at the level of miRNA-target interaction, via the regulation of mRNA binding proteins, such as RNA-binding-motif protein 38 (RBM38). This provides another level of complexity in the mediation of miRNA function by p53, and suggests another process by which p53 deletions or mutations in cancers can affect their tumor suppressive capabilities.
The importance of post-transcriptional miRNA regulation by p53 is not yet fully understood, however these studies suggest that this process is fundamentally involved in p53-mediated tumor suppression. Additionally, p53 mutations in tumors, target not only the transcriptional functions and protein-protein interactions mediated by its central DNA binding domain, but also the regulation of miRNA processing. This suggests a new mechanism by which p53 gain-of-function mutations contribute to cancer, as the effects of p53 mutations on miRNA processing will almost certainly be different to complete loss of p53 expression and this requires further evaluation.

1.3.2.3 The regulation of p53 expression by miRNAs

On the other hand, miRNAs are able to regulate p53 itself. This occurs either directly, through interactions between miRNA seed regions and the 3’UTR of TP53 mRNA, or indirectly through the down-regulation of upstream p53 regulators. Those miRNAs that bind p53 directly, function in a p53-repressive manner and include miR-125b, miR-504, miR-33, miR-380, miR-1285, miR-25 and miR-30d. These may be clinically relevant oncogenes and their overexpression has been correlated with more advanced disease and poorer survival in various tumor types. An in silico approach identified that miR-504 could bind to two sites on the p53 3’UTR and negatively regulate p53 expression. Furthermore, miR-504 over-expression impaired p53-mediated G1 cell cycle arrest and apoptosis and promoted tumorigenesis in colon cancer in vivo models.

miR-125b negatively regulates p53 in humans and zebrafish during the stress response and development. Its knockdown increased p53 protein levels and induced apoptosis in zebrafish brain and human cells, whilst its overexpression repressed endogenous p53 levels and reduced apoptosis. In colorectal cancers, elevated miR-125b expression was associated with invasion and increased tumor size and correlated with decreased survival. Interestingly, miR-125b has also been proposed to regulate many other genes in the p53 signaling pathway and mediate both apoptosis and proliferation depending on the cellular context. Le et al. (2011) used a gain- and loss-of-function screen for miR-125b targets in humans, mice, and zebrafish. These were then validated with luciferase reporter assays and a novel miRNA pull-down assay to demonstrate that miR-125b directly controls at least 20 targets in the p53 network. Amongst its targets were regulators of apoptosis such as Bak1, Igfbp3, PUMA and Prkra, and also cell-cycle regulators including cyclin C, Cdc25c and Cdkn2c. These results suggested that miR-125b shields and adjusts p53 network activity.
by modifying the activity of both apoptotic and proliferative regulators, with implications for
tumorigenesis and stem cell homeostasis.\textsuperscript{300}

A number of miRNAs indirectly control p53 activity through targeting of p53 regulators such
as MDM2, SIRT1 and HDAC1, and this regulation often occurs as part of complex
feedback/feedforward loops. miR-192, miR-194, miR-215 and miR-605 decrease MDM2
activity by direct binding, but miR-122 achieves this indirectly via the down-regulation of
cyclin G1 and subsequent inhibition of PP2A phosphatase recruitment to MDM2. Both
pathways result in reduced p53 polyadenylation and hence increased activity of this tumor
suppressor. miR-34a and miR-449 have similar seed sequences and have both been shown to
target SIRT1, leading to increased p53 acetylation and p53 induced apoptosis.\textsuperscript{242,247}

1.3.2.4 Feedback loops involved in miRNA-regulated p53 signaling

The expression and activity of p53 is under the control of miRNAs, and p53-regulated
miRNAs mediate tumor suppression and stress responses by regulating p53-signaling
pathways. They achieve this by directly targeting the stability and translation of central
mRNA components of these processes, usually via their involvement in various types of
feedforward and feedback loops that mediate the integrity, amplification, buffering and fine-
tuning of signals, and that jointly contribute to appropriate cellular reactions.\textsuperscript{242} This may
occur via simple positive or negative feedback loops, whereby a miRNA can either amplify
or suppress signaling respectively; or via more complex coherent or incoherent feedforward
loops. These are demonstrated in Figure 2.
A signaling pathway involving p53, miR-34a and SIRT1 is an example of a positive feedback loop (Figure 2a).\textsuperscript{304} SIRT1 is an NAD(+)‐dependent deacetylase able to inhibit p53 activation. In response to stress, p53 induces miR-34a expression, which increases p53 acetylation by targeting SIRT1. The resultant increase in p53 activity intensifies p53-associated tumor suppressive signaling and further augments miR-34a expression, heightening the specific effects of this miRNA on apoptosis, cell growth arrest and senescence.\textsuperscript{304} Furthermore, miR-605 creates a positive feedback loop that enables p53 to escape the p53:MDM2 negative feedback loop. In response to cellular stress, miR-605 expression is induced due to initial p53 activation and this then feeds back to increase the transactivation activity of p53 by repressing MDM2.\textsuperscript{305}

The miR-34 and miR-200 family members are involved in a complex p53‐regulated negative feedback loop that mediates and fine‐tunes p53 (and TGF‐β) associated EMT/MET and
stemness (Figure 2b).

Cellular stress activates p53, leading to the induction of miR-34 and miR-200. In addition to targeting other genes that promote stemness and EMT, miR-34 directly targets SNAIL. This has a number of effects, including a reduction in TGF-β-activated SNAIL signaling and EMT; and a decrease in SNAIL induced repression of miR-34 and miR-200 transcription.

A coherent feedforward loop, describes a system whereby a transcription factor inhibits expression of a target gene, whilst also inducing expression of a miRNA that represses the mRNA encoded by that gene. This allows for a more reliable regulation of the target gene, as its mRNA will be repressed even if the transcription factor is intermittently dysfunctional. In an incoherent feedforward loop, expression of the target gene may be stronger as there is transcriptional co-induction of a miRNA and its mRNA target by the relevant transcription factor. Others have described a slightly different type of incoherent feedforward loop termed a ‘type-2 feedforward loop’. This is a regulatory pattern in which X represses a target Z and also represses Y, another repressor of Z; and functions to accelerate responses to specific stimuli and have a buffering effect to reduce the variance of network dosage (Figure 3). Such incoherent feedforward loops have been found in the transcription factor networks of hematopoietic stem cells and hESCs, and have been shown to modulate p53-signalling and the Myc-E2F1 pathway.
The p53, miR-145 and MYC pathway exemplifies a coherent feedforward loop (Figure 2c). p53 directly represses the expression of MYC, but induces the transcription of miR-145 which targets MYC mRNA, leading to cell cycle arrest. miR-194 is involved in an incoherent feedforward loop and its actions may in fact promote oncogenesis by increasing tumor angiogenesis. This miRNA is transcriptionally induced by p53 in HCT116 cells, and targets TSP-1, an endogenous inhibitor of angiogenesis. The promoter region of THBS1, the gene that encodes TSP-1, is directly activated by p53 and therefore there is transcriptional co-induction of both miR-194 and its mRNA gene target TSP-1 by p53 (Figure 2d). miR-125b seems to function in a ‘type-2 feedforward loop’ via its targeting of multiple genes involved in p53 signaling, although it may in fact be transcriptionally down-regulated rather than induced by p53. Several studies have implicated miR-125b as an oncogene in various malignancies, due to its direct suppression of p53 and Bak1. However, others have described its function as a tumor suppressor. This can be explained by the fact that, depending on the cell type and environmental stresses, miR-125b can target genes in the p53 network that promote both apoptosis and cell proliferation and in doing so can fine-tune and protect the p53 network dosage. miRNA-target pairs are relatively poorly conserved throughout evolution and between species, but miRNA-gene network relationships are strongly conserved. These findings are consistent with the idea that miRNAs function as buffers of gene expression. It is possible that this buffering feature of miR-125b represents a general principle of miRNA regulation of p53 and other gene networks.

Rather than depending on one or a few mediators, p53 represents a central regulator which uses numerous coding and non-coding genes to achieve tumor suppression. Therefore the experimental deletion of single p53-induced protein-coding genes rarely recapitulates the loss of p53 in terms of suppressing tumor cell growth in vitro or in vivo. Furthermore, the frequent inactivation of p53-induced miRNAs in tumors suggests that their loss may promote tumorigenesis by destabilizing or disabling the p53 pathway. Accordingly, some studies have shown that replacement of single p53-induced miRNAs can prevent tumorigenesis and cancer cell proliferation in p53-knockout models, suggesting that their ability to target more than one coding gene may confer significant therapeutic potential. However, the evaluation of the effects of single p53-induced miRNAs is complicated, as several may co-target the same mRNA and this can lead to phenotypic compensations. Bandi and Vassella (2011) demonstrated that the synergistic effect of miR-15a/16 and miR-34a was required to control cell cycle arrest (but not apoptosis) in lung cancer cell lines, as their concerted action enabled...
the down-regulation of more targets than each miRNA alone, suggesting a functional link between these miRNAs.\textsuperscript{265} Therefore the collective importance of p53-associated miRNAs and their gene targets in terms of mediation of tumor suppression requires further experimental validation and the combined deletion/up-regulation of several miRNAs may be crucial to testing and understanding their relevance to p53 function. In order to achieve this, a better appreciation of the p53-associated miRNA-mRNA interactome is required. However, in mammals it is often difficult to directly predict relevant downstream miRNA targets, as perfect complementarity between miRNAs and their gene targets rarely exists. Many algorithms have been developed that predict miRNA targets (e.g. TargetScan, PicTar and miRanda),\textsuperscript{1,318,319} but the number of targets predicted for each miRNA by these programs is usually in the tens or hundreds. The identification of global miRNA targets relevant for tumor suppression by p53 is therefore a complex task. Over the coming years, this can potentially be achieved by investigating global p53-associated changes in miRNA and mRNA levels and interactions on AGO2. These data could be used to evaluate which p53-regulated miRNAs may be utilized as diagnostic, prognostic and therapeutic tools in the cancer clinic. miRNA-replacement therapies could be used to augment the expression of p53-induced miRNAs and this approach would have the advantage of not relying on the presence of functional, wild-type p53. Interestingly, the systemic delivery of miR-34a mimetics to mouse models of prostate and lung cancer have shown therapeutic benefit and pre-clinical trials involving other p53-induced miRNAs have produced encouraging results.\textsuperscript{320,321} Furthermore, the design of novel therapies to exploit p53–miRNA target networks may be possible and these could overcome difficulties such as resistance to DNA-damaging agents in tumors with aberrant p53 function. The fact that miRNAs can target multiple mRNAs involved in the p53-signaling pathways may provide further protection against the emergence of therapy-resistant tumors during systemic treatment.

1.3.3 Identifying p53-associated miRNA Targets

The mutation or dysregulation of miRNAs contributes to or causes a number of diseases including cancer, and this is predominantly due to the down-regulation or loss of repression of tumour suppressor and oncogenic mRNA targets respectively. Therefore accurate identification of miRNA targets or target networks is vital to our understanding of miRNA function in specific cellular pathways. However, until recently web-tool based
approaches and experimental techniques were limited in their ability to provide this information.

Initial techniques allowed the identification of one or a few mRNA targets of a single miRNA in a certain cell or tissue. They generally involved expression profiling of pre-specified miRNAs using RTq-PCR, followed by the use of computational prediction software to identify potential targets. Results were then validated using knockdown or over-expression of the specific miRNA followed by mRNA and protein correlations using RTq-PCR and Western blot, as well as validation of putative targets by luciferase assays and site directed mutagenesis. This was limited to identifying only a few selected targets based on prior biological data, which is suboptimal due to the fact that many genes are synchronously regulated by many miRNAs, and that most miRNAs have hundreds of gene targets. Furthermore, it is difficult to accurately predict miRNA targets using computational methods, particularly transcriptome-wide, as the relatively small number of nucleotides in the seed sites involved in target recognition often leads to a high number of false-positive predictions. However, this problem has been partially overcome by assessing the evolutionary conservation of predicted target sites, as conserved binding sites are considered more likely to be functional. To elucidate such a complex system of miRNA-mediated regulation requires extensive target identification. To this end, substantial progress has been made through the successful development of techniques such as microarray expression analyses and next generation sequencing to measure global expression changes after miRNA transfection or knockdown, coupled with proteomics or seed-enrichment analyses. Since 2009 a number of studies have used AGO RIP-chip (a form of AGO immunoprecipitation followed by reverse transcription and a microarray) or RIP-seq (a form of AGO immunoprecipitation followed by next generation sequencing), to analyze changes in gene target expression and identify specific target features that may determine target binding. Such techniques provide an indirect measure of global miRNA regulation, but the advent of new CLIP techniques, particularly involving AGO, in conjunction with large-scale sequencing has allowed more direct target identification in various species and cell lines.
1.3.3.1 Cross-linking and immunoprecipitation (CLIP)

CLIP involves the use of ultraviolet (UV)-induced crosslinking of the miRNA-target RNAs to an RNA binding protein such as AGO, followed by cell lysis and recovery of the protein of interest along with the crosslinked RNA via immunoprecipitation. The next steps involve the use of ribonuclease (RNase) treatment to reduce the RNA targets to their protein binding sites, and subsequent co-purification of the miRNA and target genes using radioactive labeling, SDS-PAGE and electroelution. This is followed by identification of the precise miRNA target sites on the mRNA via adaptor ligation, reverse transcription, library generation and finally high-throughput sequencing and computational analysis. Various different CLIP methods have been described, including high-throughput sequencing of RNA isolated by crosslinking immunoprecipitation (HITS-CLIP), individual-nucleotide resolution UV crosslinking and immunoprecipitation (iCLIP) and photoactivatable-ribonucleoside-enhanced cross-linking and immunoprecipitation (PAR-CLIP), many of which used AGO2. In contrast to other RBPs used in CLIP protocols such as Pumlio2 and IGF2BP1, which contain a specific mRNA recognition site, miRNP (e.g AGO) targets’ are specified by the interaction of the miRNA on the RISC with partially complementary sequences in the target mRNAs.

Classical cross-linking uses a UV wavelength of 254nm in the absence of a modified nucleotide. The PAR-CLIP protocol (Figure 4) uses a UV wavelength of 365nm in the presence of a modified photoreactive nucleotide such as 4-thiouridine (4SU) or 6-thioguanosine (6SG). These crosslinked modified nucleotides induce a greater crosslinking efficiency but more importantly result in the introduction of specific mutations (G-to-A for 6SG and T-to-C for 4SU) during the reverse transcription step which is an accurate indicator for quality control and also provides a means to precisely identify the individual binding sites of the selected protein(s) on their targets. The T-to-C conversion generally occurs just outside the seed region, where AGO proteins are in close proximity to the single stranded target mRNA, which suggests that the miRNA-mRNA base pairing prevents crosslinking between AGO and 4SU in the mRNA seed region. Interestingly, contrary to initial opinion, classical cross-linking in the absence of a modified nucleotide can also lead to specific mutations that can be similarly used to pinpoint binding sites, although they occur less frequently than in PAR-CLIP. Furthermore, Granneman et al (2008), using a classical
CLIP technique demonstrated that crosslinking using UV 254nm induced an increase in nucleotide deletions at U3 snoRNA RBP binding sites. 

CLIP techniques have been used successfully to identify many miRNA target sites and to better understand miRNA interactions with RNA binding proteins. Although the computational analysis is complicated, and the technique itself is open to certain biases, CLIP does provide a more accurate method for miRNA target identification than target prediction programs. However, despite demonstrating binding of the miRNP complex to mRNA targets, CLIP does not allow accurate assessment of the effects of miRNA binding on the subsequent mRNA repression, as strength of binding does not necessarily correlate with the extent of down-regulation of the target gene. For example binding affinity can be affected by a number of factors, not least the tertiary structure of the binding protein. However this can be overcome by using CLIP in conjunction with techniques such as RIP-Chip or AGO immunoprecipitation with high-throughput sequencing, that provide quantitative estimates of miRNA function, to more accurately unravel miRNA-regulated gene networks at a transcriptome level. Furthermore, Corcoran et al. (2011) describe a method of enrichment analysis of AGO PAR-CLIP data that can be used to identify those miRNAs with the strongest impact on mRNA targeting even in the absence of expression data. 

Interestingly only around 40-60% of miRNA-target interaction sites identified by CLIP are matched to the miRNA seed region indicating that the miRNA-target interaction is more complex than previously thought. Seed pairing has previously been shown to correlate strongly with target gene expression changes and is considered to be a key factor in miRNA-target interaction. 

However, a number of other determinants of miRNA targeting have been demonstrated, and they are summarized in Table 5. Using a bioinformatics approach, Wen et al. (2011) found differences in target determinants between miRNA transfection-expression-based data and CLIP-seq data. Seed-related features such as seed type and target site number were found to be stronger determinants in expression based data than CLIP data but target flanking features such as flanking region conservation were more important in the latter than the former. Features that were highly ranked for both data types were target-site accessibility and flanking AU composition. These findings suggest that determinants of miRNA binding differ from those of mRNA degradation. Others have shown in RIP-chip experiments that U frequency in entire transcripts is associated with mRNA degradation, and that miRNAs have a target preference for short 3’-UTRs. This is
contrary to earlier studies that suggested miRNAs preferentially target genes with long 3’-UTRs containing multiple miRNA binding sites, and that highly expressed transcripts are evolved to possess shorter 3’-UTRs to avoid miRNA targeting. Furthermore, Corcoran et al. (2011) found that nucleotides immediately adjacent to a seed match in a 3’-UTR were AU rich and that there was an under-representation of guanines (Gs) surrounding the interaction sites. This latter finding may be due to the use of RNase T1 in the PAR-CLIP protocol, as this enzyme cleaves next to G. This may result in some binding sites being missed if they are immediately surrounded by Gs as this would lead to the formation of reads that are too short to be included in the library step, which specifically collects reads of 30 nucleotides in length. Furthermore, CLIP methods has identified that although the majority of interaction sites on AGO are located in the 3’ UTR, there are also a number of interaction sites found in the intergenic, coding, intronic and 5’ UTR regions, although the functional significance of this is uncertain.

The initial approach designed by Hafner et al. (2010), aligned reads on the genome, using a read length of at least 20 nucleotides and allowed for one T-to-C mismatch. They then used crosslink-centered regions (CCRs), 41-nucleotide regions re-centered on the read group location with the highest percentage of T-to-C conversion events, to identify AGO interaction sites. However, Corcoran et al. (2011) used more lenient alignment parameters, allowing read lengths as short as 13 nucleotides after adapter stripping, which could contain up to two T-to-C mismatches. They then used a kernel density estimate to calculate a ‘signal-to-noise ratio’ in order determine interaction sites. The PARalyzer method was used to test miRNA binding sites across 3’ UTRs of genes known to be expressed in HEK293 cells. The signal-to-noise ratio was higher in the top expressed miRNAs but fell to below the background level for miRNAs with very low expression. However when the PARalyzer method was used to re-analyse the CCRs reported by Hafner et al. (2010), these regions were associated with a lower signal-to-noise ratio for highly expressed miRNAs, suggesting that the newer PARalyzer technique creates a higher resolution map of miRNA binding sites than the original method. It is probable that CLIP-seq protocols may preferentially select for the miRNP-target interactions with the highest binding affinities, and with the increased use of such protocols in the search to determine miRNA targets on a genome wide scale, it is important to understand the determinants of miRNA targeting to avoid selection biases in the data. It may be that predictive models need further integration into the computational analysis.
of CLIP data in order to fully understand the relevance of certain miRNA binding interactions with their target mRNAs.

Figure 4: Diagrammatic representation of an AGO PAR-CLIP experiment
4-SU: 4-thiouridine; 32P: radioactive phosphate. 4-SU: 4-thiouridine; UV: ultraviolet; PCR: polymerase chain reaction; kda: kilodaltons.
<table>
<thead>
<tr>
<th>Feature</th>
<th>Region of interest</th>
<th>Determinant</th>
</tr>
</thead>
<tbody>
<tr>
<td>3’ UTR related</td>
<td>3’UTR</td>
<td>Length$^{329}$ Relative distance to 3’UTR ends$^{329}$ Minimum distance to 3’UTR ends$^{329}$ Number of target sites$^{329}$</td>
</tr>
<tr>
<td>Target Complementarity</td>
<td>Seed</td>
<td>Seed mismatch$^{329}$ Seed type i.e. 6-mer, 7-mer etc$^{41}$ 3’ out-seed pairing$^{41}$ Target sequence alignment$^{318}$</td>
</tr>
<tr>
<td>Conservation</td>
<td>70nt</td>
<td>Flanking region conservation$^{329}$ Seed conservation$^{329}$ Entire target sequence conservation$^{329}$</td>
</tr>
<tr>
<td>Target site flanking region</td>
<td>70nt, 30nt</td>
<td>Mono/dinucleotide frequencies$^{329}$ AU content$^{41}$</td>
</tr>
<tr>
<td>Target free-energy</td>
<td>Target</td>
<td>Free-energy loss $\Delta \Delta G$ $G_{duplex}$ miRNA-target hybridization energy$^{329}$</td>
</tr>
<tr>
<td>Flanking Strand asymmetry</td>
<td>70nt</td>
<td>Base asymmetry bias eg. G vs C$^{344}$ G+U content$^{144}$ Folding-energy difference $\Delta G$ between the two strands$^{344}$</td>
</tr>
</tbody>
</table>

Table 5: Determinants of miRNA-target interaction
UTR: untranslated region; nt: nucleotides; AU: Adenylate-uridylate; G: guanine; A: Adenine; C: cytosine.

1.4 Small nucleolar RNAs

1.4.1 The structure, synthesis and function of small nucleolar RNAs

snoRNAs are a well-characterized class of ubiquitously expressed, non-coding RNAs (ncRNAs) that are 60–300 nucleotides in length. SnoRNAs are predominantly located in the nucleolus, where they classically function as guide RNAs for the post-transcriptional maturation and modification of ribosomal RNAs (rRNAs), and snRNAs involved in the spliceosome. snoRNA guide sequences hybridize specifically to their rRNA target sequence, and, via associations with proteins, form small nucleolar ribonucleoprotein complexes (snoRNPs) and execute specific rRNA modifications. Therefore, snoRNAs are crucial for ribosomal function and the effective regulation of translation and thus, unsurprisingly, are highly conserved throughout evolution. There are two major classes of snoRNAs, termed C/D box
snoRNAs and H/ACA box snoRNAs, respectively. They differ in terms of their sequence and structure, their binding partners and the nature of the post-transcriptional modifications that they induce.346,347

C/D box snoRNAs are typically 60-200 nucleotides in length and are characterized by highly conserved Box C (UGAUGA) and Box D (CUGA) domains located near to their 5’- and 3’-ends, respectively. Although the C and D boxes are located at different ends of the molecule, an interaction between the 5’- and 3’-termini causes it to fold, forming a hairpin-like structure containing a stem region in which the C and D Box elements are brought into close proximity. This region provides a binding site for core box C/D snoRNP proteins, including the methyltransferase, fibrillarin.348 Internal copies of the Box C and D elements, termed Box C’ and Box D’, display lower sequence conservation and are located nearer to the centre of the molecule.348,349 C/D box snoRNAs form snoRNPs containing a core of four proteins (fibrillarin, nucleolar protein 56 (NOP56), NOP58 and NHP2-like 1) and define the target sites for the 2’-O-ribose methylation of specific rRNA residues. The guide region, which ensures the specificity of this modification by base pairing to the target RNA, is a region of 10-20 nucleotides found immediately upstream of the D and/or D’ boxes.348

Measuring approximately 120-250 nucleotides in length, the H/ACA box snoRNAs are generally longer than C/D box snoRNAs and function as guides for the pseudouridylation of rRNA. They are folded structures consisting of two hairpins connected by a hinge region containing the H box domain, which is characterized by an ANANNA-motif (N can be any nucleotide). The ACA box, consisting of the trinucleotide, ACA, terminates the second hairpin and is located three residues upstream of the 3’ end of the molecule. H/ACA box snoRNAs form snoRNPs containing dyskerin (a pseudouridine transferase responsible for pseudouridylation of the rRNA), GAR1, NHP2 and NOP10, and the internal stem-loops of one or both of their hairpins serve as guides for the pseudouridylation of their substrates.350,351 In addition to their role in rRNA pseudouridylation, these snoRNAs are also involved in the nucleolytic processing of precursor rRNA (pre-rRNA) and in telomeric DNA synthesis.352

Small Cajal body-specific RNAs (scaRNAs) represent a class of small RNA with similar functions to the two main types of snoRNA described above. They are clustered in the small membrane-less sub-compartments of the nucleus termed the Cajal bodies, where they have
been shown to be involved in the post-transcriptional modification of snRNAs. scaRNAs are larger than the classical snoRNAs but are structurally similar, displaying the characteristic boxes of both C/D and H/ACA snoRNAs in addition to CAB boxes (motif UGAG) which function as a Cajal body localization signal.\textsuperscript{353}

In eukaryotic genomes, snoRNAs have been shown to be predominantly encoded in the introns of protein-coding host genes but some are under the control of independent promoters.\textsuperscript{354} In humans, most snoRNAs are intronic and co-transcribed with their host gene transcripts, and then processed out of the excised introns.\textsuperscript{355} However, the transcription of a minority occurs through independent RNA polymerase II or III activity in a similar manner to many miRNAs.\textsuperscript{353,355} Closely related snoRNA family members are usually encoded in different introns of the same host gene, but some host genes encode numerous unrelated snoRNAs. Although some snoRNA host genes appear to be non-protein coding, the host genes of many intronic snoRNAs are involved in nucleolar function and protein synthesis, and as such there is often an element of co-functioning.\textsuperscript{355,356} The fact that in humans, most snoRNAs are encoded in the introns of protein-coding and non-protein-coding genes (the latter generally containing short, poorly conserved open reading frames) gives rise to the assumption that these host genes act solely as cellular housekeepers via their snoRNA-encoding sequences.\textsuperscript{357,358} However, recent studies have challenged this concept and have implicated snoRNAs and their host genes in the control of oncogenesis and cell fate.\textsuperscript{359,360} The existence of a number of ‘orphan’ snoRNAs with no known rRNA targets, and the demonstration of their presence in subcellular locations other than the nucleolus,\textsuperscript{361} supports the concept that this group of small non-coding RNAs may regulate other molecules and have additional cellular functions.\textsuperscript{362} Furthermore, a number of studies suggest an evolutionary relationship between miRNAs and snoRNAs,\textsuperscript{363} and others report that mature snoRNAs may undergo further cellular processing to form smaller snoRNA-derived RNAs (sdRNAs) with miRNA-like functions.\textsuperscript{347,363-366} Additionally, snoRNA expression has been shown to be as variable as miRNA expression in human tumour samples and normalising miRNA polymerase chain reaction (PCR) expression data to these snoRNAs introduced bias in associations between miRNA and outcome.\textsuperscript{367}
1.4.2 ‘Non-housekeeping’ roles of snoRNAs

Ender and colleagues (2008) were the first to identify a human snoRNA with a miRNA-like function. They performed a RIP-seq analysis of small RNAs associated with human AGO1 and AGO2 in multiple cell lines and identified small RNAs originating from the snoRNA, ACA45, that were able to target mRNA (e.g. CDC2L6) on the RISC complex. Furthermore, they demonstrated that ACA45 processing was independent of the Drosha-containing microprocessor complex but did require DICER; but the authors acknowledge that other transporters and nucleases may also be involved. Interestingly they also showed that ACA45 is more abundant in cells in its full-length form and that only a minor portion is processed to a small miRNA-like RNA, which is consistent with the finding that ACA45 exists as a functional snoRNA that forms snoRNPs with the protein factor GAR-1. This suggests that ACA45 is transcribed and functions as a classical snoRNA in the nucleolus of human cells, but that a minor portion is also transported to the cytoplasm by an undiscovered export receptor and then, following further processing can regulate gene expression via an interaction with the RISC complex. Interestingly, recent proteomic and immunofluorescence analyses of AGO-containing RNA–protein complexes in animals have also suggested that a relationship exists between the snoRNA and RNA silencing pathways; and mass spectrometry of AGO-associated proteins in HEK-293 cells demonstrated direct interactions between AGO1 and AGO2, and Nop56 and fibrillarin, respectively, both of which are key components of snoRNPs.

Kishore et al. (2010) provided further evidence supporting a functional role for sdRNAs by introducing the concept of a miRNA-like ‘orphan’ snoRNA. HBII-52 (SNORD 115), is a C/D box snoRNA expressed from the SNURF–SNRPN locus (the loss of which is associated with Prader-willi syndrome) that shows no clear sequence complementarity with rRNAs, tRNAs or snRNAs and is not involved in their methylation. Instead HBII-52 appears to target mRNA and exhibits sequence complementarity to an alternative exon of the human serotonin receptor 2C mRNA and regulates alternative splicing of this pre-mRNA. Furthermore, the mouse homolog of HBII-52 (MBII-52) was shown to regulate five other pre-mRNAs (DPM2, TAF1, RALGPS1, PBRM1 and CRHR1) and the authors demonstrated that MBII-52 is processed into shorter RNAs lacking the snoRNA stem that are responsible for its function.
Taft et al. (2009) further demonstrated a relationship between the RNA processing/silencing machinery and sdRNAs in eukaryotes. They demonstrated the existence of sdRNAs with evolutionary conservation in size and position that were derived from the majority of snoRNA loci in animals, Arabidopsis, and fission yeast. They showed that in animals, sdRNAs derived from C/D snoRNAs had a bimodal size distribution of 17–19 nt and >27 nt and predominantly originated from the 5’ end; and that H/ACA snoRNA-derived sdRNAs were predominantly 20–24 nt in length and originated from the 3’ end. They also showed an association between sdRNAs and the RISC complex and suggested that, in some cases, DICER was required for the processing of snoRNAs to sdRNAs. Interestingly, it appears that loss of DICER1 and DGCR8 caused a more pronounced reduction in H/ACA-derived sdRNAs than C/D-derived snoRNAs, suggesting that the production of the later may occur predominantly via a different processing pathway which may account for their bimodal size distribution.

Brameier et al. (2011) performed massive parallel pyrosequencing on small RNA fractions (up to 40nt) extracted from human CD4+ T-lymphocytes. Although over 60% of sequences corresponded to known human miRNAs, the authors identified over 9000 RNAs (approximately 2% of the total reads) of 18-28nt in length that originated from snoRNAs, which they termed sno-miRNAs. Although C/D box snoRNAs, were the primary source of sno-miRNAs in this study, many also derived from H/ACA box snoRNAs. For the C/D box derived sno-miRNAs the guide strand generally contained the box C element of the snoRNA, whilst the passenger strands were usually flanked by the box D element in their 3’-end. Following testing in further cell lines the authors demonstrated that the processing of particular snoRNAs to functional sno-miRNAs occurred in a cell line specific manner (for example, U74 was processed to sno-miRNAs in Jurkat cells but not HeLa cells). Furthermore, although luciferase assays demonstrated that many of the sno-miRNAs had gene silencing capabilities similar to those of miRNAs, this also appeared to be cell line specific. For example U27-derived sno-miRNAs had gene regulatory capabilities in HeLa and Jurkat cells but not RPMI8866 cells. Additionally, a significant number of sno-miRNAs had no significant gene silencing effects in any of the tested cell lines. For example, the majority of the sno-miRNAs derived from the GAS5 gene cluster of snoRNAs had little or no gene silencing effects (e.g. U44 and U48), although U78 did do in two of the three cell lines tested.
Scott et al. (2009)\textsuperscript{363} compared the genomic locations of known miRNAs and H/ACA box snoRNAs to investigate the possibility of an evolutionary relationship between the two classes of ncRNA. They revealed an interesting overlap between specific members of these classes and suggested that some miRNAs might have evolved from snoRNA encoding genomic regions, which they confirmed by scanning reported miRNA-encoding regions for the presence of H/ACA box snoRNA features. In fact, they showed that 20 miRNA precursors exhibited significant similarity to H/ACA snoRNAs, which included molecules predicted to target known rRNA pseudouridylation sites \textit{in vivo} for which no guide snoRNA had been reported.\textsuperscript{363} Furthermore, the predicted folded structures of these 20 H/ACA snoRNA-like miRNA precursors resembled known H/ACA box snoRNAs. The genomic regions surrounding the predicted snoRNA-like miRNAs contained transposable elements, target site duplications and poly (A) tails that correspond strongly to those regions around known snoRNA retroposons.\textsuperscript{363} Furthermore, the authors demonstrated that the precursors of five H/ACA snoRNA-like miRNAs could bind to dyskerin, suggesting that they had retained some H/ACA snoRNA functionality.\textsuperscript{363} The Brameier study\textsuperscript{347} also suggested an evolutionary relationship between snoRNAs and miRNAs by identifying a number of miRNAs on miRBase that were in fact C/D box-derived sdRNAs (HBII-99b and SNORD126) or H/ACA box-derived sdRNAs (ACA34, ACA36b and HBI-61).\textsuperscript{347}

These studies clearly suggest that, in addition to their roles as regulators of ribosomal function, snoRNAs have other important functions such as the control of gene expression at the post-transcriptional level by either directly targeting mRNA or through their conversion to smaller miRNA-like sdRNAs. In view of these various functions, it is not surprising that data are accumulating supporting an important role for snoRNAs in a host of cellular pathways including those involved in development, metabolism and carcinogenesis. An imbalance in the expression of specific snoRNAs could theoretically promote or prevent carcinogenesis as the effect that this could have on rRNA modification and maturation could affect the translation and hence expression of specific tumour suppressor genes or oncogenes. Furthermore, the ability of snoRNAs or sdRNAs to directly target tumour suppressor genes or oncogenes at the mRNA level suggests another mechanism by which they could promote or prevent carcinogenesis.

A study in zebrafish showed that suppressing the expression of three snoRNAs (U26, U44 and U78), either by disrupting host gene splicing or by inhibiting snoRNA precursor
processing, reduced snoRNA-mediated methylation of rRNA targets and caused severe morphological defects and embryonic lethality. Others showed that the loss of three C/D box snoRNAs (U32a, U33 & U35a) encoded in the ribosomal protein L13a locus conferred resistance to lipotoxic and oxidative stress in vitro and prevented the propagation of oxidative stress in vivo.

Liao et al. (2010) profiled snoRNA signatures in 22 human non-small cell lung cancer (NSCLC) tissues and matched noncancerous lung tissues, as well as plasma expressions of any identified snoRNAs in 37 NSCLC patients, 26 patients with chronic obstructive pulmonary disease, and 22 healthy controls in order to develop a novel diagnostic biomarker panel for NSCLC, similar to previously described miRNA-based diagnostic platforms. Using microarray technology with qRT-PCR validation, they identified six snoRNAs (SNORD33, SNORD66, SNORD73B, SNORD76, SNORD78, and SNORA42) as being statistically differentially expressed between the paired tumour and noncancerous samples and three of these snoRNAs (SNORD33, SNORD66 and SNORD76) were also differentially expressed in the plasma of NSCLC patients and healthy controls. The same group subsequently selected SNORA42 for further investigation into its potential role in tumorigenesis of NSCLC, as its host gene is located on chromosome 1q22 which is frequently amplified in this disease. They observed genomic amplification and associated high expression of SNORA42 in lung cancer cell lines but there was no overexpression of its host gene KIAA0907. Furthermore SNORA42 knockdown in NSCLC cells inhibited tumorigenicity in vitro and in vivo, whereas SNORA42 overexpression increased cell growth and colony formation.

Dong et al. (2009) identified the snoRNA, U50, as a candidate tumour suppressor gene in human breast, suggesting that it plays a role in the development and/or progression in this malignancy. U50 is located on chromosome 6q14.3-15 which is within the 6q14-16 region frequently deleted in breast cancer. U50 underwent frequent somatic and germline deletions and transcriptional down-regulation in breast cancer cell lines as well as primary breast tumours, and functionally, re-expression of U50 inhibited colony formation in breast cancer cell lines. The same group also identified that U50 is mutated and transcriptionally down-regulated in prostate cancer and that its overexpression prevented colony formation is prostate cancer cell lines. These two studies suggest the U50 has a malfunctioning role in
the development and progression of breast and prostate cancer. Another study has also shown that down-regulation of the H/ACA box snoRNA, h5sn2, is associated with meningioma.\textsuperscript{376}

1.4.3 SnoRNAs encoded for by the growth arrest-specific transcript 5 gene

The growth arrest-specific transcript 5 gene (\textit{GAS5}), located at 1q25, is a non-protein-coding multiple snoRNA host gene comprising of 12 exons,\textsuperscript{377,378} initially discovered during screening for potential tumor suppressor genes expressed at high levels during growth arrest.\textsuperscript{12,13} In humans, it encodes ten intronic C/D box snoRNAs and two mature long non-coding RNAs (lncRNAs) that originate from alternative 5′-splice donor sites in exon 7.\textsuperscript{378} The open reading frame encoded within GAS5 exons is short and is not thought to encode a functional protein. Mapping of its 5* end demonstrates that it possesses an oligopyrimidine tract characteristic of the 5*-terminal oligopyrimidine (5*TOP) class of genes that accumulate during cell cycle arrest but are rapidly degraded by nonsense-mediated decay during cell growth. The classification of \textit{GAS5} as a 5*TOP gene offers an explanation as to why it is a growth arrest specific transcript as while the spliced \textit{GAS5} RNA is normally associated with ribosomes and rapidly degraded, during arrested cell growth it accumulates in mRNP particles. Interestingly, the only regions of conservation between mouse and human \textit{GAS5} genes are their snoRNAs and 5*-end sequences,\textsuperscript{378} suggesting that these are the most important functional components. Although \textit{GAS5} plays a role in post-transcriptional modification of ribosomal RNA through its snoRNAs, a number of recent studies have implicated this gene in other important cellular processes.\textsuperscript{359,367,379,380} The GAS5 lncRNA was shown to interact with the DNA-binding domain of the glucocorticoid receptor where it acts as a riborepressor, influencing cell survival and metabolic activities during starvation by modulating the transcriptional activity of this receptor.\textsuperscript{359} Furthermore, the same group showed in prostate cell lines, that GAS5 mRNA sequesters the androgen/androgen receptor complex and prevents its binding to target DNA sequences,\textsuperscript{359} which is likely to play an important role in modulating the effects of androgens in the prostate. GAS5 transcripts have also been shown to be important regulators of cell survival and apoptosis in human T-cells and breast and prostate cancer cell lines,\textsuperscript{379-381} and their overexpression sensitized mammalian cancer cell lines to inducers of apoptosis.\textsuperscript{379} Furthermore, reduced expression of GAS5 and/or its snoRNAs has been demonstrated in head and neck squamous cell carcinoma,\textsuperscript{367} breast cancer\textsuperscript{367,379} and glioblastoma multiforme,\textsuperscript{382} whilst over-expression of U44, U76 and U78
has been shown in NSCLC.\textsuperscript{373} The aberrant GAS5 expression demonstrated in breast and head and neck cancer was associated with poor prognosis.\textsuperscript{367} Despite these data, little is known as to the precise role of specific GAS5 snoRNAs in the pathways in which they have been implicated, and even less is accepted about the mechanisms underlying them.

**Aims and objectives**

It is known that p53 co-ordinates cell signalling pathways through regulating the expression of gene targets. This is achieved mainly through transcriptional regulation of mRNA expression but also through the ability of p53 to regulate miRNA abundance and hence fine tune target gene expression post-transcriptionally through AGO2 mediated mRNA translational repression and/or mRNA degradation. To understand this process in more detail, we aimed to comprehensively illustrate the network of miRNA-mRNA interactions that occurs in response to DNA damage using RNA-seq, AGO2 RIP-seq and PAR-CLIP-seq to determine sets of genes (RNA seq) and miRNAs (small RNA-seq) whose expression levels change in response to DNA damage and use AGO2 IP approaches to investigate the differential loading of miRNAs and mRNAs onto AGO2. Furthermore, we wished to combine the AGO2 RIP and PAR-CLIP approaches with deep sequencing to determine the exact interaction sites between the AGO2-bound miRNAs and their mRNA targets in response to double-strand break DNA damage. In addition, we wished to investigate the p53-associated miRNA/mRNA interactome at a genome wide level, and to establish the functional effects of miRNA/mRNA binding in terms of p53-regulated signaling networks. We also aimed to find previously unidentified p53-dependent post-transcriptional processes that could affect mRNA or miRNA expression or function, and in particular, whether p53 might regulate the loading of miRNAs onto AGO2 and hence control gene expression through an unexplored mechanism of miRNA regulation in response to stimuli. Furthermore, we planned to identify differential AGO2 binding sites in response to DOX treatment that might occur independently of the miRNAs regulated by the DNA damage. Combining these findings with a motif discovery analysis, we wanted to see if any particular RNA motif present on the UTR of the mRNA targets was necessary for this regulation to occur. Finally, given the previously described role of GAS5 in the regulation of apoptosis and the well documented role for p53 in the same process, we aimed to further investigate the relationship between p53 and the GAS5 snoRNAs to gain further insight into their potential role in cell survival and oncogenesis in colorectal cancer both *in vivo* and *in vitro*. 
Chapter 2: Materials and methods
## 2.1 Materials

### 2.1.1 Mammalian cell culture

Table 6: List of mammalian cell lines

<table>
<thead>
<tr>
<th>Cell Type</th>
<th>Tissue</th>
<th>Morphology</th>
<th>Tumourigenicity</th>
</tr>
</thead>
<tbody>
<tr>
<td>HCT116 wild-type</td>
<td>Colon</td>
<td>Epithelial</td>
<td>Human colorectal carcinoma cell line</td>
</tr>
<tr>
<td>HTC116 p53-knockout</td>
<td>Colon</td>
<td>Epithelial</td>
<td>Human colorectal carcinoma cell line</td>
</tr>
<tr>
<td>DLD-1</td>
<td>Colon</td>
<td>Epithelial</td>
<td>Human colorectal carcinoma cell line</td>
</tr>
<tr>
<td>DLD-1 Dicer knock-down</td>
<td>Colon</td>
<td>Epithelial</td>
<td>Human colorectal carcinoma cell line</td>
</tr>
<tr>
<td>RKO</td>
<td>Colon</td>
<td>Epithelial</td>
<td>Human colorectal carcinoma cell line</td>
</tr>
<tr>
<td>RKO Dicer knock-down</td>
<td>Colon</td>
<td>Epithelial</td>
<td>Human colorectal carcinoma cell line</td>
</tr>
</tbody>
</table>

Table 7: Normal growth media

<table>
<thead>
<tr>
<th>Cell Type</th>
<th>Media</th>
<th>Additives</th>
<th>Storage</th>
</tr>
</thead>
<tbody>
<tr>
<td>HCT116</td>
<td>McCoy’s 5A medium (modified) (Gibco®)</td>
<td>2mM Glutamine</td>
<td>4°C, used within one month</td>
</tr>
<tr>
<td></td>
<td></td>
<td>50 units/ml Penicillin</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>50μg/ml Streptomycin</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>10% fetal calf serum (FCS)</td>
<td></td>
</tr>
<tr>
<td>DLD-1</td>
<td>RPMI (Roswell Park Memorial Institute) 1640 (Sigma-Aldrich®)</td>
<td>2mM Glutamine</td>
<td>4°C, used within one month</td>
</tr>
<tr>
<td>DLD-1 Dicer knock-down</td>
<td></td>
<td>50 units/ml Penicillin</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>50μg/ml Streptomycin</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>10% fetal calf serum (FCS)</td>
<td></td>
</tr>
<tr>
<td>RKO</td>
<td>DMEM (Dulbecco’s Modified Eagle’s Medium)(Sigma-Aldrich®)</td>
<td>2mM Glutamine</td>
<td>4°C, used within one month</td>
</tr>
<tr>
<td>RKO Dicer knock-down</td>
<td></td>
<td>50 units/ml Penicillin</td>
<td></td>
</tr>
</tbody>
</table>
### 2.1.2 Buffers and Solutions

Table 8: List of buffers and reagents

<table>
<thead>
<tr>
<th>Reagent</th>
<th>Recipe</th>
<th>Storage</th>
</tr>
</thead>
<tbody>
<tr>
<td>1M DTT (Dithiothreitol)</td>
<td>1.54g of DTT in total of 10ml ddH&lt;sub&gt;2&lt;/sub&gt;O (double distilled MilliQ water)</td>
<td>100μl aliquots at -20°C</td>
</tr>
<tr>
<td>1M Tris-HCl</td>
<td>60.5g Tris in total of 500ml ddH2O and adjusted to desired pH with pure HCl</td>
<td>Room temperature</td>
</tr>
<tr>
<td>Laemmli’s buffer</td>
<td>50mM Tris-HCl pH 6.8, 15% glycerol, 0.1% (w/v) bromophenol blue, 4% SDS (BioRad)</td>
<td>Room temperature</td>
</tr>
<tr>
<td>SDS (Sodium Dodecyl Sulphate)</td>
<td>125mM Tris-HCl pH 6.8, 20% glycerol, 4% SDS, 100mM DTT, 0.04% (w/v) bromophenol blue</td>
<td>500μl aliquots at 4°C</td>
</tr>
<tr>
<td>loading buffer (2x)</td>
<td>10g SDS, 30.3g Tris, 144.1g glycine dissolved in 1l of ddH&lt;sub&gt;2&lt;/sub&gt;O</td>
<td>Room temperature</td>
</tr>
<tr>
<td>10x SDS PAGE Running buffer</td>
<td>30.3g Tris, 144.1g glycine dissolved in 1l of ddH&lt;sub&gt;2&lt;/sub&gt;O</td>
<td>Room temperature</td>
</tr>
<tr>
<td>TBS 10x (Tris-Buffered Saline)</td>
<td>24.23g Trizma HCl, 80.06 g NaCl dissolved in 1l of ddH&lt;sub&gt;2&lt;/sub&gt;O and adjusted pH to 7.6 with pure HCl</td>
<td>Room temperature</td>
</tr>
<tr>
<td>TBS 10x (Tris-Buffered Saline)</td>
<td>100ml of TBS 10x, 900ml ddH2O, 1ml Tween&lt;sup&gt;®&lt;/sup&gt; 20 (BDH)</td>
<td>Room temperature</td>
</tr>
<tr>
<td>Blocking solution</td>
<td>5% dried skimmed milk powder in TBST</td>
<td>Prepared as required</td>
</tr>
<tr>
<td>RIP Washing buffer</td>
<td>50mM Tris-HCL pH7.5 (2.5ml of 1M) 300mM NaCl (3ml of 5M) 5mM MgCl2 (2.5ml of 100mM) NP40 0.05% (25ul) Make up to 50ml with ddH2O</td>
<td>Prepared as required</td>
</tr>
<tr>
<td>RIP Lysis buffer</td>
<td>20mM Tris-HCL pH7.5 (1ml of 1M) 150mM KCL (7.5ml of 1M) 0.5% NP40 (250ul) 2mM EDTA (200ul of 500mm) 1mM NaF (500ul of 100mM) Make up to 50ml with ddH2O Just before use add to this the following for a 20 ml</td>
<td>Prepared as required</td>
</tr>
</tbody>
</table>
batch of lysis buffer: 160u/ml of RNAsin (80ul), 0.5mM DTT (100ul), Protease inh 50x (1 tab in 1ml ddh20) (400ul), Phosphotase inh 1x (2 tab in 2ml is 10x stock) (2ml)

4-Thiorudine stock
Made to a final conc of 50mM by adding 100mg to 7.69ml of ddh20

PAR-CLIP NP40 lysis buffer
50mM HEPES pH7.5 (2.5ml of 1M) 150mM KCL (7.5ml of 1M) 0.5% NP40 (250ul) 2mM EDTA (200ul of 500mM) 1mM NaF (500ul of 100mM) Make up to 50ml with ddH20
Just before use add to this the following for a 20 ml batch of lysis buffer: 0.5mM DTT(10ul of 1M), Protease inh 50x (1 tab in 1ml ddh20) (400ul)

PAR-CLIP Citrate-phosphate buffer
94mg citric acid and 184 mg Na2HPO4 in 50ml made to pH 5.0 (4.7g/l citric acid and 9.2g/l Na2HPO4 gives pH 5.0)

PAR-CLIP wash buffer
50mM HEPES-KOH pH7.5≈ 2.5ml of 1M 300mM KCL≈ 15ml of 1M 0.05% NP40≈ 25ul Make up to 50ml with ddH20
Just before use add to this the following for a 20 ml batch of IP wash buffer: 0.5mM DTT (10ul of 1M), protease inh 50x (1 tab in 1ml ddh20) (400ul)

PAR-CLIP High salt wash buffer
50mM HEPES-KOH pH7.5 (2.5ml of 1M) 500mM KCL (25ml of 1M) 0.05% NP40 (25ul) Make up to 50ml with ddH20
Just before use add to this the following for a 20 ml batch of high salt wash buffer: 0.5mM DTT (10ul of 1M), Protease inh 50x (1 tab in 1ml ddh20) (400ul)

PAR-CLIP Dephosphorylation buffer
50mM Tris-HCL pH7.9 (2.5ml of 1M) 100mM NaCl (1ml of 5M) 10mM MgCl2 (5ml of 100mM) Make up to 50ml with ddH20
Just before use add to this the following for a 20 ml batch of Dephosphorylation buffer: 1mM DTT (20ul of 1M)

PAR-CLIP Phosphatase wash buffer
50mM Tris-HCL pH7.5 (2.5ml of 1M) 50mM NaCl (0.5ml of 5M) 20mM EGTA (10ml) 0.5% NP40 (250ul) Make up to 50ml with ddH20

PAR-CLIP PNK buffer
50mM Tris-HCL pH7.5 (2.5ml of 1M) 50mM NaCl (0.5ml of 5M) 10mM MgCl2 (5ml of 100mM) Make up to 50ml with ddH20
2.1.3 Antibodies/probes for western blots & immunoprecipitation

Table 9: Primary antibodies

<table>
<thead>
<tr>
<th>Antibody</th>
<th>Dilutions for western blots</th>
<th>Company</th>
<th>Dilution buffer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monoclonal mouse p53 (DO-1)</td>
<td>1/5000</td>
<td>Santa Cruz, California, USA</td>
<td>Blocking buffer</td>
</tr>
<tr>
<td>Monoclonal rat AGO-2</td>
<td>1/5000</td>
<td>Sigma-Aldrich, Dorset, UK</td>
<td>Blocking buffer</td>
</tr>
<tr>
<td>Monoclonal mouse Dicer</td>
<td>1/2500</td>
<td>Santa Cruz, California, USA</td>
<td>Blocking buffer</td>
</tr>
<tr>
<td>Monoclonal mouse β-actin</td>
<td>1/100000</td>
<td>abcam®, Cambridge, UK</td>
<td>Blocking buffer</td>
</tr>
<tr>
<td>Monoclonal mouse E-cadherin (HECD-1)</td>
<td>1/1000</td>
<td>Takara Bio Inc., Shiga, Japan</td>
<td>Blocking buffer</td>
</tr>
<tr>
<td>Monoclonal mouse GAPDH</td>
<td>1/1000</td>
<td>abcam®, Cambridge, UK</td>
<td>Blocking buffer</td>
</tr>
</tbody>
</table>

Table 10: Secondary antibodies

<table>
<thead>
<tr>
<th>Antibody</th>
<th>Dilution</th>
<th>Company</th>
<th>Dilution buffer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Polyclonal rabbit anti-rat IgG/HRP</td>
<td>1/2500</td>
<td>Dako, Cambridge, UK</td>
<td>Blocking buffer</td>
</tr>
<tr>
<td>Polyclonal goat anti-mouse IgG/HRP</td>
<td>1/2500</td>
<td>Dako, Cambridge, UK</td>
<td>Blocking buffer</td>
</tr>
<tr>
<td>Polyclonal goat anti-rat IgG/HRP</td>
<td>1/1000</td>
<td>EMD Millipore Corporation, Billerica, MA, USA,</td>
<td>Blocking Buffer</td>
</tr>
</tbody>
</table>
2.1.4 Reagents used in the luciferase assays

Table 11: List of luciferase 3'UTRs

<table>
<thead>
<tr>
<th>Gene 3'-UTR</th>
<th>Product ID</th>
<th>Company</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cyclin D1 (CCND1)</td>
<td>S813994</td>
<td>SwitchGear Genomics, Menlo Park, CA</td>
</tr>
<tr>
<td>Pogo transposable element with ZNF domain (POGZ)</td>
<td>S810968</td>
<td>SwitchGear Genomics, Menlo Park, CA</td>
</tr>
</tbody>
</table>

Table 12: Transfection media and reagents

<table>
<thead>
<tr>
<th>Transfection Media</th>
<th>Additives</th>
<th>Transfection reagent</th>
<th>Storage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Opti-MEM® Reduced Serum Medium(Gibco®)</td>
<td>No</td>
<td>Lipofectamine® 2000 Reagent (Invitrogen, Life Technologies Ltd, Paisley, UK)</td>
<td>4°C</td>
</tr>
</tbody>
</table>

2.1.5 Primers used for reverse transcription

All the primers were designed by myself and synthesized by Sigma-Aldrich (Dorset, UK)

Table 13: Primers used for reverse transcription and their sequences

<table>
<thead>
<tr>
<th>Gene</th>
<th>Primer sequence (5' to 3')</th>
<th>Gene</th>
<th>Primer sequence (5' to 3')</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>forward AGGGCTGCTTTTAACTCTG</td>
<td>p21</td>
<td>forward GGAGACTCTCAGGGTCGAAA</td>
</tr>
<tr>
<td></td>
<td>reverse CCCCACCTTTGATTTTGAGG</td>
<td></td>
<td>reverse GGATTAGGGCTTCCCTTTTG</td>
</tr>
<tr>
<td>U44</td>
<td>forward TGATAGCAATATGCTGACTGA</td>
<td>U47</td>
<td>forward GGAGACTCTCCTCTCTTTGG</td>
</tr>
<tr>
<td></td>
<td>reverse CAGTTAGAGCTAATTAAGAC</td>
<td></td>
<td>reverse CAAATGCTGACTGATGATAG</td>
</tr>
</tbody>
</table>

2.2 Methods

2.2.1 Mammalian cell culture

2.2.1a Growing and passaging cells

Cells were cultured in 150 cm² flasks or 100-mm dishes or 6-well plates unless otherwise specified and maintained at 37°C in a humidified 5% CO2 incubator. They were routinely passaged when a confluence of ~90% was reached. During cell passage, medium was aspirated, cells were washed once with PBS solution and then trypsinised with EDTA-trypsin at 37°C for 3 minutes to allow detachment. FCS was added to inactivate the trypsin
(1:1 ratio) and cell clumps were disrupted by gentle pipetting. The cell suspension was pipetted out of the flask, transferred to a 15 mL sterile centrifuge tube and centrifuged for 4 min at 1300 rpm. After centrifugation, the supernatant was aspirated and the cell pellet was re-suspended in the appropriate volume of medium. The resulting suspension was split to the desired dilution into new flasks and fresh media was added. In all experiments cells had undergone between 4 and 20 passages.

2.2.1b Cell treatments with doxorubicin

Cells were plated in 150 mm dishes at a 50% confluence and incubated under normal growth conditions to adhere. They were then treated with doxorubicin at a concentration of 0.2ug/ml or equivalent volume of vehicle (ddh2o). After each treatment time point, dishes were placed on ice and medium was aspirated. Cells were washed twice with cold PBS, scraped and centrifuged for 5 minutes at 1300 rpm. The supernatant was removed and the cell pellet was processed for RNA and/or protein extraction.

2.2.2 Quantitative real-time Reverse Transcription-PCR

2.2.2a RNA preparation

Working under hood cells were lysed using 1ml of Trizol reagent (Invitrogen) and vigorous pipetting. Lysates were collected in eppendorf tubes and 200µl of chloroform per 1ml of Trizol reagent were added. Samples were vortexed vigorously for 15 seconds, incubated at RT for 2 to 3 minutes, and then centrifuged at 12,000 x g for 15 minutes at 2 to 8°C. Following centrifugation, each mixture separates into a lower red, phenol-chloroform phase, an interphase, and a colourless upper aqueous phase, with RNA remaining exclusively in the aqueous phase which was transferred carefully into fresh eppendorf tubes. Total RNA was precipitated from the aqueous phases by mixing with 500µl of isopropyl alcohol and incubating for 10 minutes at RT. The mixtures were stored overnight at -80 °C to allow small RNA precipitation and the RNA precipitates were then centrifuged at 12,000 x g for 10 minutes at 2 to 4°C to form a pellet on the side and bottom of the tube. The supernatants were discarded, and the RNA pellets were washed in 1ml of 75% ethanol and vortexed. Samples were centrifuged at 7,500 x g for 5 minutes at 2 to 8°C. The supernatants were removed and pellets were air-dried to remove residual ethanol. Finally the RNA pellets were re-suspended in an appropriate volume of RNAsre-free water. Subsequently, RNA concentration was measured at 260nm and 280nm wavelengths, using a NanoDrop ND-100-Spectrophotometer.
(NanoDrop Technologies). The RNA quality was determined on non-denaturing agarose gel electrophoresis: for each RNA sample isolated, two intensive bands at approximately 4.5 and 1.9 kb were observed at the transilluminator. These bands represent 28S and 18S rRNA and indicate successful RNA preparation. RNA extraction from 5 to 8 10µm FFPE sections of each tumour tissue sample was performed using the RNeasy FFPE kit (Qiagen), in accordance to manufacturer’s instructions.

2.2.2b cDNA synthesis

Reverse transcription of mature miRNAs and selected pri-miRNAs and was performed using 10ng of total RNA and the TaqMan MicroRNA Reverse Transcription Kit (Applied Biosystems). For measurement of gene or selected pri-miRNAs expression, cDNA was synthetized from 1 µg of purified Dnase-treated RNA by the Superscript III First Strand cDNA synthesis system (Invitrogen). Reverse transcription reaction were performed by incubating the samples in a 7900Ht Thermal Cycler (Applied Biosystems). After RT cycles, the cDNAs samples were placed in ice and then prepared for quantitative real-time PCR.

2.2.2c Quantitative real-time PCR

To amplify mature miRNAs and selected pri-miRNAs, for a single reaction, 1 ng of relative cDNA template was distributed in a Fast Optical 96-well reaction plate (Applied Biosystems), followed by the appropriate volume of TaqMan® Universal PCR Master Mix, No Amperase® UNG and the relative 20x Real Time TaqMan probe (both Applied Biosystems). For both gene-specific expression and selected pri-miRNAs, 10 ng of cDNA per reaction were amplified using the Power SYBR green PCR master mix (Applied Biosystems). Quantitative real-time PCR (qPCR) was performed with an ABI Prism 7900HT sequence detection system (Applied Biosystems). Data were analysed using qBasePlus software (biogazelle).

2.2.3 SDS-polyacrylamide gel electrophoresis and western blot

2.2.3a Protein extraction

Cell pellets or fresh frozen tissue samples were lysed in 30 to 60 µl of NP-40 lysis buffer + protease inhibitors cocktail solution (Roche). eppendorf tubes containing cell lysates were rotated for 15 minutes at 4°C. and then microcentrifuged at maximum speed (13,000
rpm) for 15 minutes at 4°C to allow separation of proteins from insoluble elements. The supernatants containing proteins were transferred into new eppendorf tubes and subjected to protein quantification.

2.2.3b Protein quantification

Protein concentration was calculated using the Bradford Reagent Kit (BioRad). Absorbance readings were measured at 595nm using a Beckman DU® 530 Life Science UV/Visible spectrophotometer. After data collection, the concentration of the unknown samples was determined based on standard absorbance value. The protein samples were then prepared for the SDS-polyacrilamide gel electrophoresis: 3µg/µl of protein samples were mixed with SDS Loading Buffer (5x) and boiled at 95°C for 5 minutes.

2.2.3c SDS-polyacrylamide gel electrophoresis

Both acrilamide (8-20%) resolving and 4% stacking gels were prepared manually as required. Rainbow markers (Fermentas, Life science) and the protein samples (10µl) were loaded and electrophoresis was carried out for 2 to 3 hours at 80V. Proteins were separated by SDS-polyacrilamide gel electrophoresis (SDS-PAGE) and run in a running buffer.

2.2.3d Western blots

Proteins were transferred to a Hybond C super nitrocellulose membrane (GE Healthcare) for 3 hours at 100V in transfer buffer using a Mini-PROTEAN® Tetra Cell (BioRad). The membrane was stained with Ponceau S solution (Fluka) to check if the transfer was successful. The membrane was washed once for 10 minutes with blocking buffer in order to remove the staining and then blocked with blocking buffer at RT for 1 hour. The membrane was transferred to the primary antibody solution and left at 4°C overnight. Membranes were washed three times for 15 minutes each in TBST. An IgG/HRP secondary antibody diluted in blocking solution was then added, and the membrane was incubated at room temperature for 1 hour. Membranes were washed 3 times with TBST and Enhanced Chemiluminescence (ECL) detection system (GE Healthcare) was used for visualization. The emitted fluorescence was detected using Hyperfilm ECL (GE Healthcare) on SRX-101A x-ray developer.
2.2.4 Immunoprecipitation

2.2.4a Preparation and washing of the sepharose beads

Lysis and washing buffers were prepared as per Table 8. Forty microliters of sepharose beads were used for each immunoprecipitation sample and for pre-clearing of each lysate. Beads were spun down in the centrifuge by pulsing and the supernatant was removed. They were then washed 3 times in 1ml of washing buffer and the supernatants were removed. Beads were resuspended in 1ml lysis buffer with additives and aliquoted out in equal volumes into the required number of eppendorf tubes (1 per sample).

2.2.4b Preparation of the cell lysate

Cells (e.g. HCT116 p53WT, HCT116 p53−/−, RKO or DLD-1) were plated in 150 mm dishes at a 50% confluence and incubated under normal growth conditions to adhere. They were then treated with doxorubicin at a concentration of 0.2ug/ml or equivalent volume of vehicle (ddh20) for 24 hours. The cell plates were then removed from the incubator and the medium was completely removed. The cells were then washed twice in cold PBS, scraped in 1ml of PBS, collected into falcons (1 for each point) and centrifuged at 1300 rpm for 3min. The PBS supernatant was removed and cells were re-suspended in lysis buffer with additives and passed 3 times through a syringe connected to a 25G needle and once through a 27G needle. Lysates were incubated on ice for 15mins and then centrifuged at 16000xg at 4°C for 10 minutes. Lysates were transferred into a clean falcon and pre-cleared by the addition of sepharose beads followed by rotation at 4⁰C for 2 hours. The lysates/bead mixtures were then spun down and the supernatants transferred to new falcons for each point and the beads discarded. The lysates were then filtered using using a 0.45um syringe. 10% of total lysate was removed and kept as the input samples and the remainder used for the immunoprecipitation steps.

2.2.4c Conjugation of antibodies to the beads, and the immunoprecipitation steps

10ug of anti-AGO (anti-p53 or anti–Dicer) or anti-IgG antibodies were added to the appropriate bead samples in addition to 1mg/ml of heparin and were then rotated for 2 hours at 4⁰C. Eppendorf's containing the antibody/bead mix were spun down, the supernatant was removed and the beads were washed 3 times in washing buffer and then re-suspended in lysis.
buffer. The lysates prepared previously were then divided evenly and added to the appropriate antibody/bead mixtures. The samples were rotated for 4 hours at 4°C and then spun down. 10% of the supernatants were removed and kept in eppendorfs as the flow through and the remaining supernatants were discarded. The beads then underwent a prolonged washing step using lysis and washing buffers.

2.2.4d Extraction of RNA from the immunoprecipitated samples using phenol chloroform

Samples were spun down and the supernatants were removed. Beads containing the samples, and the input and flow-through samples were reconstituted with 30ul of DNAse buffer and 1ul of DNase I and left at RT for 15min. 300ul of 2x protease buffer was added followed by 3ul of protease K (20mg/ml) and the samples were incubated at RT for 20min on a thermomixer. 300ul of phenol-chloroform was then added to each samples followed by vortexing and then centrifugation at 13,000xg for 5min at 4°C. The supernatants containing the RNA were transferred to new eppendorfs and 1/10 volume of 3M NaAC was added followed by vortexing. 2ul of glycogen carrier was added to each sample followed by 2.5 volume of 100% ethanol. Samples were left at -20°C overnight and then centrifuged at 13,000xg speed for 20min at 4°C the following day. The supernatants were discarded and the remaining pellets were washed in 1ml of 70% ethanol and centrifuged at 13,000xg for 10 minutes at 4°C. The ethanol was completely removed and the pellets were air-dried and dissolved in ddH2O. RNA concentrations were determined using the Nanodrop machine and the input and flow through samples were then run on a 1% agarose gel to check the RNA quality.

2.2.5 PARCLIP

Buffers were prepared as per Table 8. The 150mm plates containing doxorubicin or vehicle treated HCT116 p53WT and HCT116 p53-/- cells were removed from the incubator. A total of 400million cells were used for each experiment. Each plate was treated with 4-thiouridine at a final concentration of 100uM, fourteen hours prior to UV-crosslinking.

Cells were washed once in 10ml of ice PBS and the PBS was removed completely. Plates were places uncovered on ice and the cells were Irradiated with 0.15 J/cm² (of 365nm UV
light in a Stratalinker 2400. Cells were then scraped cells with a scraper in 1ml of PBS per plate, transferred to 50 ml falcon and centrifuged at 500 x g for 5mins at 4°C. The supernatant was then discarded to leave a wet cell pellet. Cell lysis and the first RNaseT1 digest were then performed as follows. Three volumes of 1x NP40 lysis buffer with additives was added to the cell pellet which was then incubated on ice for 10 minutes and centrifuged at 13000 x g for 15mins at 4°C to clear the cell lysate. Lysates were then filtered by passing through a 0.2um membrane syringe filter and collected into a 15ml falcon. Lysates were then treated with RNase T1 (Fermentas 1,000U/u) to a final concentration of 1U, incubated for 15 mins at 22°C on a heat block and then cooled on ice for 5 mins.

AGO immunoprecipitation and radiolabelling of the crosslinked RNA was then performed as follows. 10ul of Dynabeads Protein G magnetic particles (Invitrogen) per ml of cell lysate were transferred to a 1.5ml eppendorf. The eppendorf was placed on a magnetic stand for two minutes and the supernatant was then removed. While the eppendorf was still on the magnetic stand, the beads were washed twice in 1ml of citrate-phosphate buffer and then resuspended in citrate-phosphate buffer at twice the original volume of bead suspension. AGO antibody was added to the tube containing the beads at a final concentration of 0.05ug/ul and the tube was rotated at room temperature for 40 minutes to allow conjugation of the antibody to the beads. The beads were then washed twice in 1ml of citrate-phosphate buffer to remove any unbound antibody, and resuspended in citrate-phosphate buffer at twice the original volume of beads suspension. The freshly prepared antibody-conjugated magnetic beads were then added to each falcon containing the RNase T1 treated lysate at a volume of 20ul per ml of lysate. Falcon tubes were placed on a magnetic particle collector for five minutes to allow the beads to collect at the bottom and sides and the supernatant was removed and discarded. The beads in each falcon were resuspended in 1ml of IP wash buffer, transferred to clean eppendorf tubes and washed twice in 1 ml of IP wash buffer using the magnetic stand for eppendorfs. After the final wash and with the eppendorfs still on the stand the original bead volume of IP wash buffer was added to the eppendorfs and the beads were resuspended by flicking. A second RNase T1 treatment was then performed using a final concentration of 10U/ul (not 100U/ul to avoid too much digestion of the RNA. The samples were then incubated at 22°C for 15mins on a heatblock, cooled on ice for 5 minutes and washed three times in high salt wash buffer. The supernatant was removed supernatant after the final wash and the beads were resuspended in one times the original bead volume of
dephosphorylation buffer. The beads were then treated with calf intestinal alkaline phosphatase at a final concentration of 0.5 U/ul and incubated on a thermocycler at 37\(^\circ\)C for 10 mins. The beads were then washed twice in 1ml of phosphatase wash buffer, placed back on the magnetic stand and then the supernatant was removed, leaving the beads stuck to the side. The beads were then washed twice in polynucleotide kinase buffer (PNK) without DTT, the supernatant was then removed and the beads were resuspended 1x the original bead volume of PNK buffer. The RNA segments were then radiolabelled using Y-\(^{32}\)P-ATP to a final concentration of 0.5uCi/ul and T4 PNK to a final concentration of 1U/ul, followed by incubation for 30min at 37\(^\circ\)C on a thermocycler. Non-radioactive ATP was then added to at a final concentration of 100uM and the samples were incubated for another 5 minutes at 37\(^\circ\)C on a thermocycler and then washed five times in 800ul of PNK buffer without DTT. The samples were placed back on the magnetic stand, the supernatants were removed and the beads were then resuspended in 70ul of SDS-PAGE loading buffer.

SDS-PAGE and electroelution of the crosslinked RNA-protein complexes was then performed as follows. The radiolabeled samples were incubated at for 5min at 95\(^\circ\)C on a heat block to denature and release immunoprecipitated RBP and crosslinked RNA. Samples were vortexed, placed on the magnetic stand and supernatants were then transferred to a clean 1.5cm eppendorf. The 70ul of supernatant for each sample was split in half and loaded into two wells of a Novex Bis-Tris 4-12% 1mm think precast polyacrylamide gel (invitrogen) and run at 80v until the sample had run through the well and then 100v for a further two hours. The gel chamber was then dissassembled and the gel was mounted on one plate to facilitate alignment of the gel following exposure on photographic film. The gel was then wrapped in Saran wrap to avoid contamination, placed in a cassette and transferred to a dark room where a film was placed in the cassette and left at room temperature for 15-20 minutes to allow exposure. D-tube dialyzers (Dialyzer Midi Tubes) were then prepared by adding 800ul of sterile water, leaving them upright for five minutes and then removing the water. The gel bands corresponding to the expected size of AGO (approx. 100kda.) were cut for each sample and transferred to the D-tube Dialyzer Midi Tube and 800ul of 1x SDS running buffer was added to the tube. The crosslinked RNA-RBP complex was electroeluted in 1x SDS running buffer for 2h at 100V. The electric current was then reversed for two minutes to release the protein from the tube membrane. The D-tube dialyzers were then opened and the eluate was pipetted up and down five times on the inner side of membrane taking care to avoid the gel slice and not to puncture the membrane. Eluates were then transferred to a clean eppendorf
tube and centrifuged for 1 min at 14000 x g to pellet the gel residues and the supernatants were transferred to clean eppendorfs. A proteinase K digestion was then performed by adding 2x proteinase K buffer at an equal volume to the electroeluate followed by proteinase K to a final concentration of 1.2 mg/ml. Samples were then incubated for 30 min at 55°C on the thermocycler.

Recovery of cross-linked target RNA fragments was then performed as follows. 300 ul of phenol-chloroform was added to each sample which were then vortexed and centrifuged at 13,000 x g for 5 min at 4°C. The supernatant phase was transferred to a new eppendorf and 1/10th of the volume of 3 M NaAC was added. Samples were vortexed immediately and 1 ul of glycoblu was added followed by 3 volumes of 100% ethanol. Samples were then mixed and left at -20°C overnight. Following this samples were defrosted and centrifuged at maximum speed for 20 minutes at 4°C to form a pellet at the bottom of the eppendorf. The supernatants were removed and discarded and the pellets were washed in 1 ml of 70% ethanol. The samples were centrifuged at maximum speed for 10 min at 4°C, the ethanol was removed and the pellets were air dried for 5 minutes and then dissolved in 6 ul of sterile water. cDNA libraries were then prepared from these small RNA fragments.

2.2.6 Preparation of the RNA and small RNA libraries

RNA and small RNA libraries were produced using the Illumina TruSeq DNA Sample Prep Kit ® and Small RNA Sample Prep Kit ® respectively (Illumina, San Diego, USA), as per the manufacturer’s protocols. For both procedures, 200 ng of RNA were used from the immunoprecipitation samples and 4 ug from the input samples. For the RNA sequencing (RNA-seq), paired-end sequences (reads) of 100 nt in length were then generated using a HiSeq 2000 instrument (Illumina). For the small RNA-seq, single-end reads of 50 nt in length were generated using a HiSeq 2000 instrument (Illumina).

2.2.7 3′-UTR luciferase reported assays

The 3′UTRs of gene targets identified in the PAR-CLIP were cloned into pLightSwitch_3UTR GoClone vectors (SwitchGear Genomics, Menlo Park, CA) that were then used to confirm direct target binding. The day before transfections were performed, HCT-116 cells were seeded onto 24-well plates (10 × 10⁵ cells per well) in antibiotic-free media in triplicate per construct. Cells were then allowed to adhere and were grown at 37°C.
with 5% CO₂ for 24 h. Cells (80% confluent) were then co-transfected with pLightSwitch_3UTR luciferase reporters (50 ng/well) and the relevant pre-miR (100 nmol/L) or a nontargeting negative control using Lipofectamine® 2000 (Invitrogen, Carlsbad, CA). Each transfection was performed in triplicate in three independent experiments. Luciferase assays were performed using the LightSwitch Luciferase assay reagent system (SwitchGear Genomics, Menlo Park, CA). Twenty-four hours after transfection, HCT116 cells were lysed with 50µl per well of the Cell Culture Passive Lysis Buffer (5x) (Promega, Madison, WI, USA) diluted fivefold in ddH₂O and placed on an agitator for 30 minutes at a constant speed. These lysates were then transferred to an Opti-plate 96-well plate and mixed with 50µl of 1X LightSwitch assay reagent per well (SwitchGear Genomics, Menlo Park, CA). The plate was then covered to protect the reagents from the light, and left to stand for 30 minutes. Following this, the 96-well plate was sealed and the firefly luciferase activity was measured using a luminometer. Luminescence measurements were then calculated, and averages of triplicates were determined.

2.2.8 Collection, handling and RNA extraction from laser captured micro-dissected (LCM) tumour samples

With the approval of our institutional review board, tissue specimens representing normal colonic tissue and colonic adenocarcinoma were obtained immediately after surgery, cut into blocks, and then formalin fixed and embedded in paraffin. Prior to microdissection, eight 8-µm serial sections were cut (−25°C) from the same tissue block and placed onto slides (1mm) covered with polyethylene naphthalate membrane that had been heated to 180 degrees for 4 hours prior to use to remove RNAses. After mounting, the slides were allowed to dry overnight in an oven at 56°C. Deparaffinization was performed using xylene and ethanol.

Immediately after sectioning and mounting, slides were stained with Hematoxylin and Eosin (H & E) in the following way. Slides were dipped 5 times in RNase-free distilled water and were stained for 1-2 minutes in Mayer’s Hematoxylin solution (SIGMA, #MHS-32). They were then rinsed for 1 minute in diethylpyrocarbonate (DEPC)-treated tap water and stained for 10 seconds in Eosin Y (SIGMA, #HT110-2-32), followed by a quick increasing ethanol series (70%, 96%, 100%). They were then microdissected using the PALM Laser MicroBeam system (P.A.L.M. Microlaser Technologies GmbH, Bernried, Germany). This system
employs a high-energy laser beam to microdissect along a precise, predefined line and to catapult the tissue areas-of-interest from the slide into a collection cap of an eppendorf tube containing 6 µm solution [20 µL 0.5 M EDTA, pH 8.0, 2000 µL 1 M Tris, pH 8.0, 50 µL Igepal® Ca 630 (Sigma-Aldrich, St. Louis, MO, USA), and diethylpyrocarbonate (DEPC)-treated double-distilled water]. A total area of 200,000 µm² was micro-dissected from each slide. Figure 5 shows images of various stages of the microdissection process.

Figure 5: Photographs of tissue blocks on microdissection slides at various stages of the procedure. N= normal colon, T= tumour, A= adenoma.1: Hematoxylin and eosin staining; 2: areas to be microdissected have been marked; 3: slide following microdissection of marked areas; 4: microdissected tissue following catapulting onto the inside of the eppendorf caps.
RNA was then extracted from the micro-dissected samples in the following way. A volume of 150μl of Buffer PKD (Qiagen) was added to the eppendorf tubes containing the LCM elements in the adhesive cap. 10μl of Proteinase K was then added and the tubes were inverted to allow contact between the reagents and the samples on the adhesive surface. The samples were then digested by placing the tubes upside down in an incubator at 56°C overnight, after which they were vortexed and heated at 80°C for 15 minutes in a heat block and then incubated on ice for 3 minutes. Following this, 16 μl of DNase Booster Buffer (Qiagen) and 10 μl of DNase I stock solution (Qiagen) were added and the samples were mixed gently by inverting the tubes which were then centrifuged briefly to collect any residual liquid from the sides of the tubes. The samples were then incubated at room temperature for 15 minutes, and the lysates were transferred to a new 1.5 ml microcentrifuge tube. 320μl of RBC (Qiagen) were added to adjust the binding conditions and the lysates were mixed thoroughly, prior to the addition of 720μl of 100% ethanol to the samples which were then mixed well by pipetting. 700μl of the sample were transferred to an RNeasy MinElute spin column (Qiagen) placed in a 2 ml collection tube and centrifuged for 15 seconds at ≥10000 rpm. The flow-throughs were discarded and this step was repeated until the entire sample had passed through the RNeasy MinElute spin column. Following this, 500μl of Buffer RPE (Qiagen) were added to the RNeasy MinElute spin column which was then centrifuged for 15 seconds at ≥10000 rpm to wash the spin column membranes. The flow-throughs were then discarded and the RNeasy MinElute spin columns were then placed in a new 2 ml collection tube. The spin columns were then centrifuged at full speed for 5 minutes with their lids opened and the flow-throughs in the collections tubes were discarded. The RNeasy MinElute spin columns were placed into new 1.5 ml collection tubes and 30μl of RNase-free water were added directly to the spin column membranes. The lids of the collection tubes were closed gently and the tubes were centrifuged for 1 minute at full speed to elute the RNA. The RNA solution was then stored at -20°C or used directly for reverse transcription.

Prior to use, the quality and quantity of isolated total RNA were assessed using the RNA 6000 Pico LabChip kit and the 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). 550ul of RNA 6000 Pico gel matrix were placed on a spin filter, centrifuged at maximum speed for 10 s, and divided in 65-μL aliquots. 1 μl of RNA 6000 Pico dye was added to each aliquot, and the mixture was mixed by vortexing and centrifuged for ten minutes at maximum speed. Using the priming solution, each aliquot was filled with gel-dye
mix, conditioning solution, marker, 1 µL RNA 6000 ladder, and RNA samples, and was then vortex mixed for 1 min before being run on the Bioanalyzer. Both the 28s/18s rRNA subunit ratio and the RNA integrity number (RIN) were used to evaluate RNA integrity.

2.2.9 Collection, handling and RNA extraction from macro-dissected colorectal tissue

Specimens of normal, adenomatous and malignant colorectal tissue were obtained from individuals undergoing colorectal surgery or colonoscopy between 2011 and 2013 at St Mary’s Hospital, London, UK. These were immediately macrodissected at the time of surgery, placed directly in RNALater stabilization solution (Qiagen, Hilden, Germany), stored at room temperature for 2–3 hours, and then frozen at −80°C. This was done to prevent any degradation and to ensure high-quality RNA for the subsequent molecular profiling. H&E staining was used for histological confirmation of cancer and to determine the cellularity of representative sections. A specialist colorectal pathologist reviewed the slides, and tissue for RNA isolation was verified to contain ≥60% neoplastic cells. Written informed consent was obtained from all patients, and ethical approval was provided by the hospital's research ethics committee.

Fresh tissue stored in RNALater (Qiagen) was crushed in liquid nitrogen and the subsequent powder lysed in Trizol (Invitrogen, Paisley, UK) reagent. RNA isolation was carried out according to manufacturer's instructions, as described above in section 2.2.2 (‘RNA preparation’). The quality of isolated RNA was analyzed on the Agilent 2001 Bioanalyzer (Agilent Technologies Inc., Santa Clara, CA).

2.2.10 Next generation sequencing and bioinformatic analysis of results

2.2.10a Description of samples and experimental design

RNA-sequencing and PAR-CLIP methods were used to profile the abundance of small RNAs (3 biological replicates), poly-adenylated RNAs (2 biological replicates) and AGO-bound RNAs (3 biological replicates, with 3 technical replicates each) in total RNA or IPed RNA samples from human colon cancer HCT116 cell lines from human colon cancer
HCT116 cells with either wild-type or knockout TP53. Table 14 describes the different samples assayed.

<table>
<thead>
<tr>
<th>Doxorubicin treatment</th>
<th>-</th>
<th>+</th>
<th>-</th>
<th>+</th>
</tr>
</thead>
<tbody>
<tr>
<td>p53 status</td>
<td>-/-</td>
<td>+/-</td>
<td>+/-</td>
<td></td>
</tr>
<tr>
<td>Total RNA</td>
<td>InV, InD, IwV, IwD</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AGO-IPed RNA</td>
<td>AnV, AnD, AwV, AwD</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IgG-IPed RNA</td>
<td>GnV, GnD, GwV, GwD</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 14: Summary of Samples.**
Cell line: HCT116 colorectal carcinoma human cell line. Abbreviations: -/-: p53 knockout cell lines; +/-: p53 wilt-type cell lines. I: input or total RNA; A: AGO-IPed RNA; G: IgG-IPed RNA; n: p53 knockout; w: p53 wild type; D: doxorubicin treated; V: vehicle treated.

**2.2.10b Processing of small RNA sequencing data**

*Read preprocessing.* Raw sequences were cleaned from any sequence derived from the sequencing library adapters using cutadapt version 1.2.1. *Read mapping.* Clipped reads were mapped to the human genome (hg19) by using bowtie\(^{384}\) (version 1.0.0). *Read annotation to human miRNAs.* Mirdeep\(^2\)\(^{385}\) and miRNA sequence annotations from miRBase\(^{386}\) (release 20) were used to identify and quantify known and novel mature miRNAs. Figure 6 summarizes the aforementioned steps.

The miRDeep analysis quantified the abundance of 2578 mature human miRNA sequences annotated in miRBase\(^{386}\) - release 20. The analysis was then focused on identifying miRNAs expressed in at least one sample condition (i.e. any combination of p53+/+ or p53/-/ cells; doxorubicin or control treatment; Total RNA or AGO-IPed sample). The dataset was first restricted to miRNAs having at least six raw read counts in at least two samples [N=1322 mature miRNAs, including redundant identifiers owed to multiple stem-loops yielding the same mature sequence; N=1185 unique mature miRNAs]. The DESeq Bioconductor package\(^{387}\) was used to estimate library size factors and normalize read counts across samples.
accordingly. Figure 7 shows the unsupervised clustering of samples in the miRNA dataset based on normalized read counts for 1322 selected miRNAs.

**Figure 6: Outline of small RNA sequencing data processing.**
Cutadapt was used to clip adapter sequences from the raw small RNA sequencing data. Different modules from miRDeep together with Bowtie were then applied to collapse clipped reads from the different samples, map them to the human genome (hg19) and quantify both known and novel human mature miRNAs. Known human miRNA sequence annotations were taken from miRBase.
Figure 7: Unsupervised clustering of miRNA samples based on normalized read counts. The 1322 miRNAs included in the count dataset used to cluster the samples were selected by setting an abundance threshold of at least six raw read counts in at least 2 samples. Normalization of read counts across samples was performed by using the DESeq Bioconductor package based on estimated library sizes. Abbreviations. Clustering of samples is based on euclidean distance of log2-transformed read counts. A tri-letter code specifies the sample type: 1st letter: Total (T) or AGO-IPed (A) RNA; 2nd letter: p53+/+ (w) or p53−/− (n) cells; 3rd letter: doxorubicin (D) or vehicle (V) treatment. The subsequent part of each sample name indicates the biological replicate (r2,r3,r4 for the 1st,2nd and 3rd replicates, respectively)
2.2.10c Analysis of miRNA modulation following DNA-damage

To find miRNAs showing differential expression following induction of DNA-damage following doxorubicin treatment, mean log-fold changes (mean of 3 biological replicates) were computed for all miRNAs [N=1322] for the following four comparisons of doxorubicin-treated vs. vehicle-treated cells: [i.] Total RNA samples from p53+/+ cells (TwD.vs.TwV); [ii.] Total RNA samples from p53-/− cells (TnD.vs.TnV); [iii.] AGO-IPed RNA samples from p53+/+ cells (AwD.vs.AwV); [iv.] AGO-IPed RNA samples from p53-/− cells (AnD.vs.AnV). Figure 8 reports histograms for all of the above contrasts.

A moderated t-test for paired samples was used to evaluate significance of the observed modulation. To reduce the number of tests, we restricted the statistical significance analysis to top-abundant and modulated miRNAs in each given comparison by setting arbitrary thresholds for minimum average abundance (log2-scale) and absolute log-fold change to 5 and 0.35, respectively. For comparisons involving AGO-IPed samples, we also required miRNAs to match the above abundance threshold in the total RNA samples. The outlined selection yielded 227, 209, 175 and 135 tests for significant modulation for miRNAs in the TwD.vs.TwV, TnD-vs.TnV, AwD-vs.AwV and AnD-vs.AnV groups, respectively (Figures 9 & 10). A Benjamini-Hochberg adjusted p-value of the moderated t-test < 0.05 was regarded as significant.

Lists of regulated miRNAs were further grouped based on the dependancy of their modulation on TP53-status and they were defined as follows:

miRNAs regulated in a TP53-dependent manner: [i] a significant modulation (i.e.: adjusted p-value< 0.05) in TP53+/+ doxorubicin-treated samples compared to untreated controls, but not in TP53-/− background; [ii] a mean log fold-change in TP53+/+ doxorubicin-treated vs untreated samples exceeding that of TP53-/− background; [iii] either an absolute mean log fold-change in TP53-/− cells below the minimum change threshold (0.35) or a significantly different extent of modulation in TP53+/+ compared to TP53-/− cells (i.e.: t-test p-value < 0.05 when comparing log fold-change values of doxorubicin-treated vs untreated samples between TP53 wild-type and TP53 knockout HCT116 cells).
miRNAs regulated in a TP53-independent manner: [i] a significant modulation (i.e.: adjusted p-value < 0.05) in TP53+/+ doxorubicin-treated samples compared to untreated controls; [ii] same as before, but for TP53-/- cells; [iii] comparable degree of modulation in TP53+/+ and TP53-/- cells (i.e.: no significant difference when comparing log fold-change values of doxorubicin-treated vs untreated samples between TP53 wild-type and TP53 knockout HCT116 cells).

Subsets of differentially expressed miRNAs falling in each of the above-listed subcategories (i.e.: TP53-dependent and TP53-independent) were individually evaluated for the Total RNA and AGO-IPed samples, as well as for the upregulated and down-regulated miRNAs.

2.2.10d Processing of long RNA sequencing data

Read preprocessing. The quality of processed reads was assayed using FastQC - version 0.10.1. Raw sequences from paired-end Illumina sequencing were processed by using TopHat and Cufflinks and according to the published protocol. Briefly, paired-end reads were mapped to the human genome (hg19) by using TopHat - version 2.0.9. Individual transcripts obtained from mapped reads were assembled into a reference transcriptome and quantified by using Cufflinks - version 2.1.1. Figure 11 summarizes the above procedure.

Analysis of gene modulation following DNA-damage. Gene expression levels were compared between doxorubicin-treated and control samples (two biological replicates for each experiment) by using Cuffdiff, a separate program included in the Cufflinks suite of software. The differential expression analysis was repeated for the same comparison groups between doxorubicin-treated and vehicle-treated cells as those listed for the miRNA analysis (see Analysis of miRNA modulation following DNA-damage). Genes included in the Cufflinks transcriptome assembly (N=29888 Cufflinks identifiers) were filtered to remove those with very low abundance. In particular, we required genes to have a Cufflinks status flag of "OK" and an FPKM value above zero for Cufflinks identifiers to be considered in the next analyses. The above yielded filtered datasets of size 13165, 12528, 13090 and 13039 for the TwD.vs.TwV, TnD.vs.TnV, AwD.vs.AwV and AnD.vs.AnV groups, respectively. Figure 12 reports histograms for all of the above comparator groups based on the log fold change values of Cufflinks identifiers retained in the filtered datasets. We next focused the analysis on
Cufflinks identifiers annotated to a gene symbol and made these datasets unique based on gene annotations (i.e. for each gene symbol, we retained only its first occurrence in the dataset sorted by decreasing absolute value of mean log fold change). For each comparison, we identified lists of modulated genes based on an absolute z-score-transformed log fold change value equal to or exceeding 1.5 (Figure 13).

Figure 8: Histograms demonstrating miRNA mean fold changes in doxorubicin-treated vs. untreated samples.
Mean fold changes were computed for all miRNAs (N=1322) across 3 biological replicates in the indicated compared groups of doxorubicin-treated vs. vehicle-treated samples. Abbreviations. TwD.vs.TwV: doxorubicin-treated vs. untreated total RNA samples from p53+/+ cells; TnD.vs.TnV: as described above, but for p53-/- cells. AwD.vs.AwV: doxorubicin-treated vs. untreated AGO-IPed RNA samples from p53+/+ cells; AnD.vs.AnV: as described above, but for p53-/- cells.
Figure 9: Scatterplots of Log Fold-Change vs Average Expression for TP53+/+ cells
Mean log fold-changes for doxorubicin-treated vs untreated samples are depicted as a function of their mean expression (mean of 3 biological replicates) for all miRNAs (top and bottom left panels) or selected miRNAs based on minimum abundance and change thresholds (top and bottom right panels). Upper panels: Total-RNA samples; Bottom panels: AGO-IPed samples. Applied filtering thresholds: average expression at least equal to 5 (log2-scale) for samples included in the indicated contrast and - for AGO-IPed samples - also in Total RNA samples; absolute log fold-change in doxorubicin-treated vs untreated samples at least equal to 0.35. Differentially expressed miRNAs: Benjamini-Hochberg adjusted p-value < 0.05. Abbreviations. TwD.vs.TwV: doxorubicin-treated vs. untreated total RNA samples from TP53+/+ cells; AwD.vs.AwV: doxorubicin-treated vs. untreated AGO-IPed RNA samples from TP53+/+ cells.
Figure 10: Scatterplots of Log Fold-Change vs Average Expression for TP53-/- cells
Mean log fold-changes for doxorubicin-treated vs untreated samples are depicted as a function of their mean expression (mean of 3 biological replicates) for all miRNAs (top and bottom left panels) or selected miRNAs based on minimum abundance and change thresholds (top and bottom right panels). Upper panels: Total-RNA samples; Bottom panels: AGO-IPed samples. Applied filtering thresholds: average expression at least equal to 5 (log2-scale) for samples included in the indicated contrast and for AGO-IPed samples - also in Total RNA samples; absolute log fold-change in doxorubicin-treated vs untreated samples at least equal to 0.35. Differentially expressed miRNAs: Benjamini-Hochberg adjusted pvalue < 0.05. Abbreviations. TnD.vs.TnV: doxorubicin-treated vs. untreated total RNA samples from TP53-/- cells; AnD.vs.AnV: doxorubicin-treated vs. untreated AGO-IPed RNA samples from TP53-/+ cells.
**Figure 11: Outline of long RNA sequencing data processing.**

FastQC\(^{387}\) was used to check quality of the RNA sequencing data. Mapping of the reads to the human genome (hg19), transcriptome assembly and gene quantification were performed by using TopHat and Cufflinks according to a published protocol\(^{389}\). Analysis of differentially expressed genes following the induction of DNA-damage through doxorubicin treatment was assayed by using CuffDiff\(^{27}\) for the following contrasts: [i.] Total RNA samples from TP53\(^{+/+}\) cells; [ii.] Total RNA samples from TP53\(^{-/-}\) cells; [iii.] AGO-IPed RNA samples from TP53\(^{+/+}\) cells; [iv.] AGO-IPed RNA samples from TP53\(^{-/-}\) cells.

<table>
<thead>
<tr>
<th>Step</th>
<th>Command Line</th>
</tr>
</thead>
<tbody>
<tr>
<td>FastQC (control read quality)</td>
<td><code>fastqc</code> &lt;br&gt;<code>perl fastqc -a format fastq R1.fastq R2.fastq</code></td>
</tr>
<tr>
<td>TopHat (map spliced reads)</td>
<td><code>tophat --transcriptome-index refFlat_hg19 --no-mixed -p 4 -g hg19 R1.fastq R2.fastq</code></td>
</tr>
<tr>
<td>Cufflinks (assemble transcriptome)</td>
<td><code>cufflinks -p 4 -o cufflinks_output accepted_hits.bam</code> &lt;br&gt;<code>cuffmerge -g refFlat.gtf -s hg19.fa -p 4 assemblies.txt</code></td>
</tr>
<tr>
<td>Cuffdiff (differential analysis)</td>
<td><code>cuffdiff -b hg19.fa -p 4 -L doxo,vehicle -u merged.gtf rep1_doxo.bam rep2_doxo.bam rep1_vehicle.bam rep2_vehicle.bam</code></td>
</tr>
</tbody>
</table>
Figure 12: Histograms of genes mean fold changes in doxorubicin-treated vs. untreated samples.
Mean fold changes for Cufflinks gene identifiers retained in the filtered datasets across 2 biological replicates in the indicated contrasts (and gene dataset size) of doxorubicin-treated vs. vehicle-treated samples. Abbreviations. TwD.vs.TwV: doxorubicin-treated vs. untreated total RNA samples from TP53+/+ cells; TnD.vs.TnV: same as before, but for TP53/- cells. AwD.vs.AwV: doxorubicin-treated vs. untreated AGO-IPed RNA samples from TP53+/+ cells; AnD.vs.AnV: same as before, but for TP53/- cells.
Figure 13: Scatterplots of genes mean fold changes in doxorubicin-treated vs. untreated samples as a function of mean gene abundance.

Scaled mean fold changes (z-scores) for Cufflinks gene identifiers retained in the filtered datasets across 2 biological replicates in the indicated contrasts (and gene dataset size) of doxorubicin-treated vs. vehicle-treated samples. Abbreviations. TwD.vs.TwV: doxorubicin-treated vs. untreated total RNA samples from TP53+/+ cells; TnD.vs.TnV: same as before, but for TP53--/- cells. AwD.vs.AwV: doxorubicin-treated vs. untreated AGO-IPed RNA samples from TP53+/+ cells; AnD.vs.AnV: same as before, but for TP53--/-
2.2.10e Seed enrichment analysis

A miRNA is functional when loaded on the RNA-induced silencing complex (RISC), which it directs to target mRNAs via sequence-specific patterns of base pairing with partially complementary sites in their 3’UTR. RISC targets are largely dictated by complementarity between the seed region (nucleotides 2-8 at the 5’-end of the mature miRNA sequence) of the loaded miRNA and one or more sites in the mRNA 3’UTR. This principle of seed primacy is utilised by the most popular target prediction algorithms.\textsuperscript{1,318,324} To evaluate whether the observed remodelling of AGO2-bound mRNAs following induction of DNA damage could have been attributed to an association with the modulated miRNAs, we used human target predictions from the Targetscan algorithm\textsuperscript{1} and tested each AGO2-enriched (and AGO2-depleted) miRNA for enriched proportion of their predicted targets among AGO2-enriched (and AGO-depleted) transcripts. We limited the database of miRNA target predictions downloaded from the targetscan website (http://www.targetscan.org/) - release 6.1 - to the union set of gene symbols included in our filtered gene datasets (background set, N=11375). Gene redundancy in the prediction database was removed by retaining only the first occurrence of each gene symbol after sorting the database by decreasing 3’UTR length.

2.2.10f Processing of AGO2 PAR-CLIP data

Read preprocessing. We used cutadapt- version 1.2.1 -to remove adapter sequences from raw PAR-CLIP data. In addition to adapter trimming, we also used cutadapt to remove low-quality ends of reads and discard trimmed reads shorter than 13 nucleotides. FastQC- version 0.10.1 - was used to check read quality before and after trimming. Trimmed reads were mapped against the human genome (hg19) allowing up to 2 mismatches by using Bowtie\textsuperscript{384} and requiring mapped locations to be reported only for reads with, at most, 10 different genomic hits and limited to the optimal mismatch-stratum for each read. Mapped reads were checked for the distribution of mismatches to verify the expected primacy of T>C (and its reverse A>G) mutations (Figure 14), which are introduced by the PAR-CLIP method at sites of crosslinking.\textsuperscript{332} Bowtie output was purged of reads mapping in repetitive sequence regions identified by RepeatMasker (http://www.repeatmasker.org/) and whose specific locations were downloaded from the UCSC genome browser.\textsuperscript{392} PARalyzer\textsuperscript{333} was used to group mapped reads and identify ”clusters” (i.e.: AGO2 binding sites) based on T-to-C conversion
locations and read density. According to PARalyzer algorithm guidelines, input mapped reads obtained from Bowtie were further refined to retain only reads mapping uniquely to the genome and containing 0, 1 or 2 T>C (or A>G) mismatches. Figure 15 summarizes the above procedure. PARalyzer clusters were annotated to RefSeq transcripts whose genomic coordinates were downloaded from the UCSC genome browser. If a cluster mapped to a genomic location that could be annotated to multiple transcripts, it was assigned based on the following order of preference: 3’UTR, miRNA precursor, coding sequence, 5’UTR, intron, intergenic.

Figure 14: Distribution of mismatches to the human genome for PAR-CLIP mapped reads
PAR-CLIP method favor T-to-C transitions at sites of crosslinking. Mapped reads were checked for the distribution of mismatches with respect to the human genome to verify the expected primacy of T>C (and its reverse A>G) before refinement of the dataset of mapped reads to be retained for further analyses (i.e. removal of reads containing mismatches to the genome other than T>C and A>G). Abbreviations. AwD,AwV: doxorubicin-treated (AwD) and untreated (AwV) AGO-IPed RNA samples from TP53+/+ cells; AnD,AnV: same as before, but from TP53-/- cells; br[1-3]: biological replicate 1, 2 or 3.
Figure 15: Outline of PAR-CLIP data processing
Cutadapt was used to clip the adapter sequence from PAR-CLIP sequencing data. Cutadapt was also used to remove low-quality ends of reads and discard trimmed reads shorter than 13 nucleotides. FastQC was used to check read quality before and after trimming. Reads were mapped to the human genome (hg19) by using Bowtie. AGO binding sites were identified based on a combination of T-to-C conversions and read density by using PARalyzer.

Analysis of AGO2 remodelling based on PAR-CLIP data. AGO2 binding sites across different samples and replicates, as identified by the PARalyzer run, located in 3’UTRs were merged using scripts from the bedtools suite to obtain a reference set of sites amenable to differential abundance analysis between DOX-treated and untreated AGO-IPed samples. Merged AGO2 binding sites were then normalized across samples based on library size using DESeq and tested based on the negative binomial distribution as implemented in the same package (function nbinomTest) to infer differential signals. Test p-values were adjusted for multiple comparisons based on the Benjamini and Hochber method and adjusted p-values < 0.05 used to select significant features. Post-transcriptional regulation of miRNAs is largely (if not completely) mediated by sequence complementarity between the mature miRNA seed (i.e.: 6-8 nucleotides at the 5’-end of the mature miRNA sequence, starting at position 1 or 2) and short sites in the mRNA 3’UTR. The dataset of 3’UTR AGO2 binding sites, as identified by the PARalyzer run were mapped to miRNA seeds. We limited the mapping to highly and/or differentially expressed miRNAs identified in our small RNA-sequencing
experiment. If an AGO2 binding site (i.e.: PARalyzer cluster) mapped to multiple miRNA seeds in our cohort, we assigned the cluster to the miRNA whose seed-complementary site occurred closest to the maximum T-to-C conversion position in the cluster. This selection was based on previous reports that suggested that the highest enrichment of miRNA complementary regions occurred at positions immediately downstream from the predominant site of CLIP-induced transition within clusters of sequence reads. All possible 6-, 7- and 8-mers substrings spanning positions 1-8 of the mature miRNA sequence (seed sequences) for miRNAs were collected based on sequence annotations from miRBase. So obtained miRNA-mRNA interaction networks involved in the response to DNA damage were then visualized by using the Cytoscape software.

Mapping of PAR-CLIP raw data to sample type, individual biological replicate (N=3) and technical replicate (N=3 for each biological replicate) was retrieved based on the samples barcode included in the sequencing read header and according to the following mapping scheme:

1. file:JK 2 1 ATCACG L002 R1 001.fastq; des:AwD biorep1 tecrep1; barcode:ATCACG
2. file:JK 2 2 CGATGT L002 R1 001.fastq; des:AwV biorep1 tecrep1; barcode:CGATGT
3. file:JK 2 3 TTAGGC L002 R1 001.fastq; des:AnD biorep1 tecrep1; barcode:TTAGGC
4. file:JK 2 4 TGACCA L002 R1 001.fastq; des:AnV biorep1 tecrep1; barcode:TGACCA
5. file:JK 2 5 ACAGTG L002 R1 001.fastq; des:AwD biorep2 tecrep1; barcode:ACAGTG
6. file:JK 2 6 GCCAAT L002 R1 001.fastq; des:AwV biorep2 tecrep1; barcode:GCCAAT
7. file:JK 2 7 CAGATC L002 R1 001.fastq; des:AnD biorep2 tecrep1; barcode:CAGATC
8. file:JK 2 8 ACTTGA L002 R1 001.fastq; des:AnV biorep2 tecrep1; barcode:ACTTGA
9. file:JK 2 9 ACTGAT L002 R1 001.fastq; des:AwD biorep3 tecrep1; barcode:ACTGAT
10. file:JK 2 10 CACCGG L002 R1 001.fastq; des:AwV biorep3 tecrep1; barcode:CACCGG
11. file:JK 2 11 CACGAT L002 R1 001.fastq; des:AnD biorep3 tecrep1; barcode:CACGAT
12. file:JK 2 12 CACTCA L002 R1 001.fastq; des:AnV biorep3 tecrep1; barcode:CACTCA
13. file:JK 3 1 ATACCG L003 R1 001.fastq; des:AwD biorep1 tecrep2; barcode:ATACCG
14. file:JK 3 2 CGATGT L003 R1 001.fastq; des:AwV biorep1 tecrep2; barcode:CGATGT
15. file:JK 3 3 TTAGGC L003 R1 001.fastq; des:AnD biorep1 tecrep2; barcode:TTAGGC
16. file:JK 3 4 TGACCA L003 R1 001.fastq; des:AnV biorep1 tecrep2; barcode:TGACCA
17. file:JK 3 5 ACAGTG L003 R1 001.fastq; des:AwD biorep2 tecrep2; barcode:ACAGTG
18. file:JK 3 6 GCCAAT L003 R1 001.fastq; des:AwV biorep2 tecrep2; barcode:GCCAAT
19. file:JK 3 7 CAGATC L003 R1 001.fastq; des:AnD biorep2 tecrep2; barcode:CAGATC
20. file:JK 3 8 ACTTGA L003 R1 001.fastq; des:AnV biorep2 tecrep2; barcode:ACTTGA
21. file:JK 3 9 ACTGAT L003 R1 001.fastq; des:AwD biorep3 tecrep2; barcode:ACTGAT
22. file:JK 3 10 CACCGG L003 R1 001.fastq; des:AwV biorep3 tecrep2; barcode:CACCGG
23. file:JK 3 11 CACGAT L003 R1 001.fastq; des:AnD biorep3 tecrep2; barcode:CACGAT
24. file:JK 3 12 CACTCA L003 R1 001.fastq; des:AnV biorep3 tecrep2; barcode:CACTCA
25. file:JK 4 1 ATCACG L004 R1 001.fastq; des:AwD biorep1 tecrep3; barcode:ATCACG
26. file:JK 4 2 CGATGT L004 R1 001.fastq; des:AwV biorep1 tecrep3; barcode:CGATGT
27. file:JK 4 3 TTAGGC L004 R1 001.fastq; des:AnD biorep1 tecrep3; barcode:TTAGGC
28. file:JK 4 4 TGACCA L004 R1 001.fastq; des:AnV biorep1 tecrep3; barcode:TGACCA
29. file:JK 4 5 ACAGTG L004 R1 001.fastq; des:AwD biorep2 tecrep3; barcode:ACAGTG
30. file:JK 4 6 GCCAAT L004 R1 001.fastq; des:AwV biorep2 tecrep3; barcode:GCCAAT
31. file:JK 4 7 CAGATC L004 R1 001.fastq; des:AnD biorep2 tecrep3; barcode:CAGATC
32. file:JK 4 8 ACTTGA L004 R1 001.fastq; des:AnV biorep2 tecrep3; barcode:ACTTGA
33. file:JK 4 9 ACTGAT L004 R1 001.fastq; des:AwD biorep3 tecrep3; barcode:ACTGAT
34. file:JK 4 10 CACCGG L004 R1 001.fastq; des:AwV biorep3 tecrep3; barcode:CACCGG
35. file:JK 4 11 CACGAT L004 R1 001.fastq; des:AnD biorep3 tecrep3; barcode:CACGAT
36. file:JK 4 12 CACTCA L004 R1 001.fastq; des:AnV biorep3 tecrep3; barcode:CACTCA

Chronology of scripts used for parclip data analysis:

1. parclip pipeline.sh
2. filtra repeatmasker reads.sh
3. filtra reads MM e gHITS.sh
4. run paralyzer.sh
5. annotate clusters to genome.sh
Chapter 3: Results
3.1 The post-transcriptional regulation of gene expression by p53 is mediated by its ability to regulate miRNA abundance, control miRNA loading onto AGO2 and remodel the miRNA-mRNA interaction network through novel mechanisms

p53 is one of the most frequently mutated tumor suppressors and regulates gene expression predominantly through its function as a transcription factor, but also to a lesser extent, through non-transcriptional processes.\(^\text{166}\) It regulates the expression of protein-coding genes and non-coding RNAs involved in numerous cellular signaling pathways\(^\text{168}\) and is fundamental in coordinating the cellular response to stresses, such as oncogene activation and DNA damage. Mutations or inactivation in p53 are the most frequent abnormalities observed in cancer cells\(^\text{169}\). Such mutations have significant implications in terms of tumor growth and response to current treatment modalities. Many chemotherapeutic agents, in particular anthracyclines and platinum-containing compounds, as well as radiotherapy, induce cancer cell death or senescence via DNA damage and the subsequent activation of p53-mediated signalling pathways. Resistance to such treatments can occur due to mutations in components of these networks. A better understanding of the role of the various elements within these pathways and their interactions with each other may allow improvements to current anticancer treatments and the design of novel therapies. Refining our knowledge of the role of miRNAs in p53-signalling networks and a better understanding of the genes they target is crucial to achieving this.

To date, most studies designed to identify p53-regulated mRNAs or miRNAs have utilized techniques such as RT-qPCR or microarray platforms which, although informative, do not allow the most detailed analysis at the transcriptome wide level. Notably, microarrays suffer from a fundamental ‘design bias’, in that they only return results from those regions for which probes have been constructed and consequently are only as good as the databases from which they are devised. The list of known miRNAs on the database, miRBase, is continually updated and therefore even the most up-to-date array platform will not be able to detect every known miRNA. Conversely, RNA-seq provides a tool by which the whole transcriptome can be studied without any prior knowledge of it and therefore enables the detection and analysis of novel transcripts. The global identification of the targets modulated by the p53-regulated miRNAs is lacking and needs further evaluation to enable a better understanding of the p53 regulatory network. Furthermore, there are no data indicating a possible regulatory
mechanism that controls the interaction between miRNAs and AGO proteins in response to treatments (such as DNA damaging agents), or any mechanism that could either induce (through an increase in AGO loading) or reduce (through a decrease in AGO loading) their activity. Studies have shown that p53 can enhance the post-transcriptional maturation of specific miRNAs via its interaction with DDX5, a component of the DROSHA/DGCR8 complex, and can also influence miRNA target gene selection at the level of the miRNA-target interaction, via the regulation of RBM38 RNA binding protein, however further understanding of how p53 may regulate miRNA function post-transcriptionally is also lacking.

It is known that p53 co-ordinates cell signalling pathways through regulating the expression of gene targets. This is achieved mainly through transcriptional regulation of mRNA expression but also through the ability of p53 to regulate miRNA abundance and hence fine tune target gene expression post-transcriptionally through AGO2 mediated mRNA translational repression and/or mRNA degradation. To understand this process in more detail, we aimed to comprehensively illustrate the network of miRNA-mRNA interactions that occurs in response to DNA damage using RNA-seq, AGO2 RIP-seq and PAR-CLIP-seq to determine sets of genes (RNA seq) and miRNAs (small RNA-seq) whose expression levels change in response to DNA damage and use AGO2 IP approaches to investigate the differential loading of miRNAs and mRNAs onto AGO2. Furthermore, we wished to combine the AGO2 RIP and PAR-CLIP approaches with deep sequencing to determine the exact interaction sites between the AGO2-bound miRNAs and their mRNA targets in response to double-strand break DNA damage.

We induced double-strand break DNA damage in p53-wild type and -null cells using the chemotherapeutic agent doxorubicin (DOX) and performed an RNA-seq and AGO2 RIP-seq (Figure 16) to identify novel miRNAs and mRNAs regulated by the p53-dependent and p53-independent responses to DNA damage and which miRNA (if any) and mRNA transcripts were differentially loaded onto AGO2. This sequencing approach gave us the potential ability to identify previously unidentified p53-dependent post-transcriptional processes that could affect mRNA or miRNA expression or function, and in particular, whether p53 might regulate the loading of miRNAs onto AGO2 and hence control gene expression through an unexplored mechanism of miRNA regulation in response to stimuli. In addition, we used an AGO2 PAR-CLIP experiment to investigate the p53-associated miRNA/mRNA interactome at a genome wide level, and to establish the functional effects of miRNA/mRNA binding in
terms of p53-regulated signaling networks. We hoped that this approach would enable us to identify which coding genes were regulated by the miRNAs that our RIP-seq had identified as being induced or repressed following DNA damage, and to characterize the sites to which miRNAs guide the RISC complex to bind to its target transcripts. Furthermore, we planned to identify differential AGO2 binding sites in response to DOX treatment in wild type and p53-null cell lines that occurred independently of the miRNAs regulated by the DNA damage. Combining these findings with a motif discovery analysis, we wanted to see if any particular RNA motif present on the UTR of the mRNA targets was necessary for this regulation to occur. In this way we wished to determine whether p53 might also affect AGO2 binding on the mRNA targets, hence remodeling the AGO2-mRNA interaction network following DNA damage.

Figure 16: Summary of the cell treatments, sample processing and sequencing of samples used
3.1.1 A small RNA-seq approach identifies novel miRNAs that are transcriptionally regulated following DNA damage in a p53-dependent and -independent manner

Using the total RNA samples produced from treating HCT116 p53<sup>WT</sup> and p53<sup>null</sup> cells with DOX or vehicle control, we performed a small RNA-seq analysis to identify miRNAs whose relative abundance changed following DOX-induced double-strand DNA damage. Fold changes were calculated by directly comparing miRNA levels between the vehicle-treated and DOX-treated cell lysates. Furthermore, by comparing changes in miRNA abundance between p53<sup>WT</sup> and p53<sup>null</sup> cells we were able to determine whether such differential expression occurred in a p53-dependent or p53-independent manner. We considered any positive or negative fold change with a threshold of 0.35 (log2) or greater and a p-value < 0.05 corrected for multiple testing (Benjamini correction), as significant. Of the 1322 miRNAs expressed in our samples, we identified 34 miRNAs that were significantly upregulated and 26 miRNAs that were downregulated in a p53-dependent manner (Figure 17A). Conversely, we found 21 miRNAs that were significantly upregulated and 26 that were down-regulated in a p53-independent manner (Figure 17B).

Firstly, we aimed to validate our approach by confirming that our results were consistent with previous studies focusing on p53-regulated miRNAs. We detected a significant upregulation in the relative expression levels of miR-34a, miR-23a, miR-143, miR-107, miR-149 and the miR-200 family (Figure 17A), and down-regulation in miR-100 (Figure 17B), which validated our assays because the expression of all of these miRNAs has been shown previously to change in response to DNA damage in various cell lines. However, we also found that the relative abundance of previously undescribed miRNAs also changed and thus identified a novel group of miRNAs involved in the DNA damage response. For example miR-1247-5p, miR-486-5p and miR-22 were upregulated in a p53-dependent manner and miR-7974, the miR-548 family and miR-1255a were down-regulated in a p53-dependent manner (Figure 17A). Furthermore, we found that the abundance of a number of miRNAs changed in a p53-independent manner which suggested that the expression of these miRNAs was under the control of alternative transcription factors following DNA damage. The p53 homologues, p63 and p73 may be the transcription factors responsible for the changes in miRNA abundance induced by DNA damage in the absence of p53 but this requires further validation. Interestingly, for the first time we show that miR-3065-5p was the most induced
miRNA, increasing by nearly 3-fold (log2) in both p53\textsuperscript{WT} and p53\textsuperscript{null} cells suggesting that it is particularly important in the p53-independent response to DNA damage. This indicates that RNA-seq approaches can be used to identify new important miRNAs that were not yet discovered and annotated when previous similar studies were performed.

As a further validation of our results, we performed an RT-qPCR analysis to measure the differential expression of a group of identified miRNAs in RNA extracted from HCT116 cells treated with DOX or vehicle. The RT-qPCR results were consistent with our small RNA-seq analysis. In particular, we validated that miR-3065-5p increased significantly with identical fold change in response to DNA damage in both p53\textsuperscript{WT} and p53\textsuperscript{null} HCT116 cells (Figure 19A).
Figure 17: A combined RNA-seq and RIP-seq approach identifies miRNAs showing differential expression and/or AGO2 loading following DOX-induced DNA damage.

A & B, A small RNA-seq approach identifies a subset of known and novel miRNAs that are significantly up- or down-regulated in a (A) p53-dependent or (B) p53-independent manner. C & D, A combined RNA-seq and AGO2 RIP-seq identifies a group of miRNAs that demonstrate differential loading onto AGO2 in a p53-dependent manner. E, Venn diagrams showing those miRNAs that demonstrate consistent expressional fold-changes in both the RNA-seq and RIP-seq analyses. Top left: miRNAs up-regulated in a p53-dependent manner; top right: miRNAs up-regulated in a p53-independent manner; bottom left: miRNAs down-regulated in a p53-dependent manner; miRNAs down-regulated in a p53-independent manner. miRNAs demonstrating a positive or negative fold change with a threshold of 0.35 (log2) or greater and a p-value < 0.05 corrected for multiple testing (Benjamini correction), were considered as significant. Abbreviations A: AGO-IPed RNA; T: input or total RNA; G: IgG-IPed RNA; w: p53 wild type; n: p53 knockout; D: doxorubicin treated; V: vehicle treated.
Figure 18: A combined small RNA-seq and AGO2-RIP seq approach confirms that the loading of members of the let-7 family onto AGO2 is regulated in a p53-dependent manner following DOX-induced DNA damage.

Abbreviations A: AGO-IPed RNA; T: input or total RNA; G: IgG-IPed RNA; w: p53 wild type; n: p53 knockout; D: doxorubicin treated; V: vehicle treated.
**Figure 19:** RT-qPCR and Western blotting validate our small RNA-seq and RIP-seq analyses and demonstrate that the AGO2 loading affect is dependent on the presence of wild-type p53 and is not the result of changes in AGO2 abundance.

A, An RT-qPCR analysis comparing fold-changes in miR-3065-5p expression in total RNA or IP samples following an AGO2 IP in DOX or vehicle treated HCT116 p53<sup>WT</sup> and p53<sup>null</sup> cell lines. B, An RT-qPCR analysis comparing fold-changes in miR-148-5p expression in total RNA or IP samples following an AGO2 IP in DOX or vehicle treated HCT116 p53<sup>WT</sup> and p53<sup>null</sup> cell lines. C & D, An RT-qPCR analysis demonstrating that the loading of let-7 family members onto AGO2 is significantly increased in p53<sup>WT</sup> (C) but not p53<sup>null</sup> (D) cells following DOX treatment. E, An RT-qPCR analysis comparing fold changes in let-7 and mir-34a (control) expression following an AGO2 IP in DLD1 and RKO cell lines treated with DOX which suggests that the increased loading of let-7 onto AGO2 requires the presence wild-type p53. F, An RT-qPCR analysis demonstrating the fold changes in miR-34a (control) and let-7 family members at 1h, 3h, 12h, and 24h following DOX treatment in HCT116 p53<sup>WT</sup> cells. G, A Western blot analysis showing the relative expression levels of AGO2 protein in HCT 116 p53<sup>WT</sup> and p53<sup>null</sup> cells following DOX or vehicle treatment. Dotted lines in RT-qPCR analyses signify a positive or negative fold change of 0.35 (log2).

**Abbreviations**

A: AGO-IPed RNA; T: input or total RNA; G: IgG-IPed RNA; w: p53 wild type; n: p53 knockout; D: doxorubicin treated; V: vehicle treated.
3.1.2 A small RIP-seq approach demonstrates that DNA damage induces a previously undescribed mechanism that controls miRNA function post-transcriptionally by regulating the loading of a subset of miRNAs onto AGO2 in a p53-dependent manner

Next we focused on identifying novel mechanisms involved in the post-transcriptional regulation of miRNA function and gene targeting in response to DNA damage. We performed an AGO2 small RIP-seq analysis using lysates from HCT116 p53\textsuperscript{WT} and p53\textsuperscript{null} cells following treatment with DOX or vehicle, to determine whether DNA damage affected miRNA loading onto AGO2 in either a p53-dependent or p53-independent manner. Subsequently, we compared differences in the DNA-damage induced modulation of miRNA expression between the total RNA samples used in our small RNA-seq analysis and the AGO2-IP samples used in our RIP-seq analysis. This allowed us to differentiate those miRNAs whose differential expression in the RIP-seq analysis was due to transcriptional or processing effects and those whose differential expression was related to an increase or decrease in their loading onto AGO2.

Following DNA damage, the expression of certain miRNAs changed in a similar manner in both the total RNA and AGO2-IP samples. These miRNAs are represented in orange in Figure 17C and are located in the top right or lower left sections of the graph. They predominantly consist of those miRNAs shown by ourselves and others to be transcriptionally induced by p53 following DNA damage (e.g. miR-34a, the miR-200 family and miR-143). Strikingly, we also found differences in miRNA expression changes between the total RNA and AGO2-IP samples (Figure 17C). We demonstrated an increased expression of certain miRNAs (e.g. let-7 family, miR-30a and miR-96) in the AGO2-IP samples that was not detectable in the total RNA samples indicating increased loading of these miRNAs onto AGO2 in response to DOX (Figure 17C, top central portion coloured red). Furthermore, we found a reduction in the expression of other miRNAs (e.g. miR-26b, miR-142 and miR-548e) in the AGO2-IP samples that was not demonstrable in the total RNA samples indicating decreased loading of this subset of miRNAs onto AGO2 in response to DOX (Figure 17C, lower central portion coloured green). To our knowledge this represents the first finding that indicates that the loading of miRNAs onto the RISC complex can be affected by a particular stimulus. Interestingly, these differences were only detected in the p53\textsuperscript{WT} cells and not the p53\textsuperscript{null} cells (Figure 17D) and taken together these results suggest
that DNA damage can induce a p53-dependent mechanism capable of regulating the loading of specific miRNAs onto AGO2 making them more or less active in regulating their gene targets. This demonstrates that p53 can control the abundance of miRNAs on AGO2 via changes in cellular miRNA levels through the regulation of miRNA transcription and processing, but also by a previously undescribed mechanism that controls the loading of certain miRNAs onto AGO2.

3.1.3 AGO2 loading of certain miRNAs, including the let-7 family, is regulated in a p53-dependent manner following DNA damage and this process is selective to those cell lines expressing wild-type rather than mutated p53

Having identified by RNA-seq that DNA damage induced a p53-dependent effect on the loading of specific miRNAs onto AGO2 in HCT116 cell lines, we aimed to validate our findings using RT-qPCR to measure miRNAs derived from AGO2 in lysates resulting from these and additional cell lines. We were particularly interested in the let-7 miRNAs because AGO2 loading of various members of this family (let-7c-5p, let-7i-5p, let-7b-5p) increased in response to DNA damage in our RIP-seq analysis (Figure 17C and Figure 18). We treated 3 different colorectal cancer cell lines (HCT116 p53\textsuperscript{WT}, RKO and DLD1) with DOX or vehicle and performed an AGO2-IP from the lysates, followed by RT-qPCR to measure miRNA expression. We selected these cell lines to enable us to detect whether the differential loading on AGO2 was specific to cells containing wild-type p53 (HCT 116, RKO cells) rather than mutant p53 (DLD1). Again, we considered any positive or negative fold changes with a threshold of 0.35 (log2) or greater as significant.

Firstly, we used RT-qPCR to measure the expression levels of the let-7 family members in total RNA or AGO-IP samples derived from HCT116 p53\textsuperscript{WT} or p53\textsuperscript{null} cells treated with DOX or vehicle. In the AGO-IP samples, we found a significant increase in the levels of let-7c-5p, let-7i-5p, let-7e-5p, let-7f-5p, let-7b-5p, let-7g-5p and miR-98-5p but not let-7a-5p or let-7d-5p (Figure 19C). However, we did not detect any significant increases in the expression of these miRNAs in the total RNA samples (Figure 19C). Furthermore, there was no significant difference in the expression levels of these let-7 family members in the AGO-IP or total RNA samples derived from the HCT116 p53\textsuperscript{null} cells treated with DOX (Figure 19D). These results further validated our hypothesis that DNA damage induced a p53-dependent regulation on the loading of specific miRNAs, including the let-7 family, onto
AGO2. The fact that, when compared to our small RIP-seq analysis, RT-qPCR identified additional let-7 family members that were loaded more onto AGO2 following DNA damage is likely to be due to subtle differences between the two techniques.

We further validated our results using RT-qPCR to measure miRNAs whose loading onto AGO2 decreased in a p53-dependent manner following DNA damage. We found that miR-148-5p (Figure 19B) and miR-142-5p (data not shown) expression levels were significantly reduced in AGO-IP samples from HCT116 p53WT cells treated with DOX but there were no significant changes in the levels of these miRNAs in the AGO-IP samples obtained from p53null cells or the total RNA samples derived from either cell line. Again, these results are consistent with our RIP-seq data and confirm that a p53-dependent reduction in loading of certain miRNAs onto AGO2 occurs after DNA damage.

Interestingly, our RT-qPCR analysis also demonstrated that the p53-dependent regulation of miRNA loading onto AGO2 was specific to cells that expressed wild-type p53 and was not detectable in cell lines that expressed mutant p53 (Figure 19E). RKO are poorly differentiated cell lines derived from colorectal cancer and, like HCT116 p53WT cells, they contain wild-type p53. Conversely, DLD1 cells are derived from a Duke’s stage C colorectal cancer and express p53 that has a missense C > T mutation at position 241 that results in a Ser > Phe substitution. We compared miRNA expression changes between AGO2-IP samples and total RNA samples derived from DOX or vehicle treated p53 wild-type RKO and p53 mutated DLD1 cells. We found that DNA damage induced miRNA fold changes in samples derived from RKO cells that were suggestive of an AGO2 loading effect and which were consistent with our results in HCT116 p53WT cells described above (Figure 19E). However, there was no evidence of differential loading of the identified miRNAs onto AGO2 in samples derived from DLD1 cells indicating that differential AGO2 loading requires p53 to be active (Figure 19E). More specifically, although DOX treatment did not affect let-7c-5p or let-7i-5p levels in total RNA samples derived from DLD1 or RKO cell lines, it did induce an increase in the expression of both miRNAs in AGO-IP samples derived from RKO cells but not DLD1 cells (Figure 19E). miR-34a was used here as a control and interestingly, DOX treatment increased the levels of this miRNA in the AGO-IP and total RNA samples derived from both cell lines (Figure 19E). This is consistent with our previous results that demonstrated that p53 regulated miR-34a transcriptionally but did not affect its loading onto AGO2.
3.1.4 Increased loading of the let-7 family onto AGO2 following DOX-induced DNA damage is not related to an early transcriptional effect

Our results outlined above derived from experiments in which DNA damage was induced by exposing cells to DOX treatment for 24 hours. At this time-point, let-7 expression levels in p53WT cells were increased in the AGO-IP samples obtained from DOX treated cells when compared to vehicle controls, but no significant expression changes were seen in the total RNA samples. This led to the conclusion that DNA damage induced a p53-dependent increase in let-7 loading onto AGO2 and that these changes were not simply related to an increase in let-7 transcription. However it is possible that the differences demonstrated in let-7 expression levels in the AGO-IP samples compared to the total RNA samples were due to a delayed effect caused by an earlier and transient increase in let-7 transcription and/or maturation. For example, if an increase in let-7 transcription occurred after 3 hours of DOX treatment, this increase may not be detectable in the total RNA samples by 24h but may still be seen in the AGO-IP samples due to the time taken for the miRNA to be processed and then loaded onto AGO2, which would lead to a false interpretation of the results.

We therefore performed a timecourse experiment to detect whether DOX treatment did in fact induce let-7 transcription in cells exposed for less than 24 hours. We subjected HCT116 p53WT cells to DOX or vehicle treatment for 3, 6, 12 and 24 hours and used RT-qPCR to measure let-7 expressions levels in total RNA extracted from the cell lysates. We did not detect any significant induction in the expression of the relevant let-7 family members at any time-point (Figure 19F) but there was an increase in miR-34a (used as a positive control) at 24 hours. These results provided further support to the hypothesis that AGO2 loading of specific members of the let-7 family is increased in a p53-dependent manner following DNA damage.

3.1.5 The differential loading of identified miRNAs onto AGO2 following DOX-induced DNA damage is not due to changes in AGO2 expression

We wished to confirm that the changes demonstrated in the loading of miRNAs onto AGO2 following DNA damage could not simply be explained by changes in AGO2 expression. We performed a Western blot analysis using protein extracted from HCT116 p53WT and p53null cells treated with DOX or vehicle. We found that DNA damage did not cause a significant p53-dependent or -independent change in AGO2 expression (Figure
19G), which suggests that a variation in the abundance of this protein was not responsible for the changes seen in miRNA loading.

3.1.6 Combining our small RIP-seq and small RNA-seq analyses validates a group of miRNAs that are regulated in a p53-dependent or p53-independent manner following DNA damage

We wished to further investigate the role of the miRNAs that we had identified as being regulated in response to DNA damage, and to identify their gene targets. To ensure that we selected these miRNAs as accurately as possible we combined data from our small RNA-seq analysis and small RIP-seq analysis and constructed venn diagrams to identify those miRNAs that demonstrated consistent expression fold-changes between both experiments (Figure 17E). For example, to select miRNAs upregulated in a p53-dependent manner, we identified those shown to be upregulated in Figure 17A that were also upregulated in both the total RNA (TwD/TwV) and AGO-IP (AwD/AwV) samples illustrated in Figure 17C (coloured orange in the top right section of the graph). To select miRNAs upregulated in a p53-independent manner, we identified those shown to be upregulated in Figure 17B that were also upregulated in the total RNA (TwD/TwV) and AGO-IP (AwD/AwV) samples shown in Figure 17D (coloured orange in top right section of the graph). In doing so, we were able to use two different assays to confirm which miRNAs were regulated in a p53-dependent or –independent manner following DNA damage.

Of the 34 miRNAs that were significantly upregulated in a p53-dependent manner in our small RNA-seq analysis (Figure 17A), 10 were also significantly upregulated in our small RIP-seq analysis (Figure 17E) and were selected for further investigation. Of the 26 miRNAs that were down-regulated in a p53-dependent manner in our small RNA-seq analysis, only 2 were significantly down-regulated in our small RIP-seq analysis (Figure 17E). Of the miRNAs regulated in a p53-independent manner, 6 of the 21 that were significantly upregulated and 3 of the 26 that were down-regulated in the small RNA-seq analysis showed consistent fold changes in the small RIP-seq analysis and were selected for further investigation (Figure 17E).
3.1.7 An RNA-seq approach followed by an extensive bioinformatic analysis identifies genes whose abundance is regulated in a p53-dependent manner following DNA damage and illustrates their function within cellular signaling networks

Having used an RNA-seq approach to better understand DNA damage-induced modulation of miRNA expression, we wished to use the same method to gain further insight into changes in mRNA expression. Firstly, we aimed to determine p53-dependent genome wide changes in mRNA expression that occurred following DNA damage. We assessed changes in mRNA expression using total RNA samples obtained from HCT116 p53\textsuperscript{WT} and p53\textsuperscript{null} cells treated with DOX for 24 hours, and performed the experiment in duplicate to achieve better accuracy. We identified over 12,000 genes in the total RNA samples from HCT116 p53\textsuperscript{WT} cell lines, and Figures 12 and 13 (in the methods section) show the mean fold changes of these genes between DOX-treated and vehicle treated samples. Furthermore, the heat map in Figure 20 illustrates genes relevant to the DNA damage response that were found to be up- or downregulated in p53-dependent manner following DOX treatment. Next, we produced a Pearson’s plot that showed a good correlation between biological replicates in terms of p53-dependent up- & down-regulated genes (Figure 21, Pearson’s Correlation= 0.71). We analysed the cellular functions of the identified genes using the DAVID online database for gene ontology (GO) annotation and pathway enrichment analysis, and performed KEGG pathways analyses to demonstrate the validity of our methods.\textsuperscript{398,399} Web-based bioinformatics databases such as DAVID\textsuperscript{399} and GO\textsuperscript{398} are able to process large lists of genes giving outputs with biological meaning, such as the cellular pathways enriched for a subset of genes. Both GO-term and KEGG analyses apply the Fischer’s exact test for enrichment but GO-term categorises genes by function or biological role, whereas KEGG gives information about the cellular pathways in which genes are involved. Notably, this analysis specifically implicated DNA damage-induced p53-dependent upregulated genes in the positive regulation of apoptosis, cell death, cellular proliferation and the DNA damage response (Figure 22A), as well as the co-ordination of pathways involved in p53-signalling, apoptosis and cancer (Figure 22B). Conversely, the analysis specifically implicated DNA damage-induced p53-dependent down-regulated genes in controlling the cell cycle, mitosis and cell division (Figure 23A), as well as the regulation of pathways involved the cell cycle and DNA replication (Figure 23B).
Figure 20: Heat map illustrating the differential expression of a number of genes relevant to the DNA damage response that were found to be differentially expressed between DOX and vehicle treated p53$^{WT}$ and p53$^{null}$ HCT116 cells in either the total RNA or AGO IP samples.

Green: up-regulated; Red: down-regulated. TwD: DOX treated total RNA samples from p53$^{WT}$ cells; TwV: vehicle treated total RNA samples from p53$^{WT}$ cells; TnD and TnV: same as before, but for p53$^{null}$ cells; AwD: DOX treated AGO IP samples from p53$^{WT}$ cells; AwV: vehicle treated AGO IP samples from p53$^{WT}$ cells; AnD and AnV: same as before, but for p53$^{null}$ cells.
3.1.8 An RNA-seq approach followed by an in depth bioinformatic analysis identifies genes whose abundance is regulated in a p53-independent manner following DNA damage and illustrates their function within cellular signaling networks

We used the same RNA-seq approach to determine p53-independent changes in mRNA expression that occurred following DNA damage on a genome wide scale. We identified over 12,000 in the total RNA samples derived from HCT116 p53null cell lines and a number of important p53-independently up- and down-regulated genes are illustrated on the heat map in Figure 20. Again, a Pearson’s plot that showed a good correlation between the biological replicates in terms of p53-dependently up- & down-regulated genes (Figure 21, Pearson Correlation= 0.9). A DAVID analysis specifically implicated DNA damage-induced p53-independent upregulated genes in the co-ordination of angiogenesis, cell adhesion and cell motility (Figure 24). Conversely, the analysis implicated DNA damage-induced p53-independent down-regulated genes in controlling the cell cycle, cell division and transcription (Figure 25).

Figure 21: Pearson’s correlation plots demonstrate consistency between replicates in terms of gene expression changes in the RNA-seq experiments.
A & B, Gene expression (mRNA) levels, as determined by RNA-seq, were measured in RNA derived from experimental replicates of HCT116 p53WT (A) or p53null (B) cells treated with DOX or vehicle and correlations between replicates were calculated using the Pearson’s correlation formula.
Figure 22: Top GO-term and KEGG analyses of genes identified as being up-regulated in a p53-dependent manner following DOX treatment in the RNA-seq.

A) GO term analysis showing the function of genes identified as being up-regulated in a p53-dependent manner in the RNA-seq. B) KEGG pathway analysis illustrating the cellular signalling pathways in which p53-dependently up-regulated genes identified in the RNA-seq are involved.
Figure 23: Top GO-term and KEGG analyses of genes identified as being down-regulated in a p53-dependent manner following DOX treatment in the RNA-seq.

A) GO term analysis showing the function of genes identified as being down-regulated in a p53-dependent manner in the RNA-seq. B) KEGG pathway analysis illustrating the cellular signalling pathways in which p53-dependently down-regulated genes identified in the RNA-seq are involved.
Figure 24: Top GO-term and KEGG analyses of genes identified as being up-regulated in a p53-independent manner following DOX treatment in the RNA-seq
A) GO term analysis showing the function of genes identified as being up-regulated in a p53-independent manner in the RNA-seq. B) KEGG pathway analysis illustrating the cellular signalling pathways in which p53-independently up-regulated genes identified in the RNA-seq are involved.
3.1.9 p53 post-transcriptionally modulates the activity of a subgroup of genes involved in the DNA damage response through differential targeting by AGO2 and this process involves p53-regulated miRNAs

Having used an RNA-seq approach to identify genes that showed differential expression at the mRNA level following DNA damage, we wished to assess whether any of
these mRNAs were specifically targeted post-transcriptionally by miRNAs regulated by DNA damage. Firstly, we made lists of genes that were differentially loaded onto AGO2 in the RIP-seq. By assessing p53-dependent AGO2 enriched genes we aimed to determine those whose expression at the protein level might be suppressed by increased miRNA-mediated post-transcriptional regulation of the corresponding mRNA following DNA damage. Furthermore, by assessing p53-dependent AGO2 depleted genes we aimed to determine those whose activity might be increased through a reduction in miRNA-mediated post-transcriptional regulation of the corresponding mRNA following DNA damage. We identified 1366 of such AGO2-enriched genes (i.e. log fold change $\geq \log_2(1.5)$ in DOX-treated AGO2-IP samples compared to vehicle-treated AGO2-IP samples) which included genes such as RET, cyclin-dependent kinase 6 (CDK6) and fibroblast growth factor 18 (FGF18). Furthermore, we found 1543 AGO2-depleted genes (i.e. log fold change $\leq \log_2(1.5)$ in DOX-treated AGO2-IP samples compared to vehicle-treated AGO2-IP samples). Interestingly, a number of the AGO2-enriched genes and AGO2-depleted genes were downregulated and upregulated respectively in the RNA-seq or their expression did not change.

These findings suggested that the activity of this subset of genes that were enriched or depleted on AGO2 following DNA damage were specifically controlled by miRNAs that were regulated in a p53-dependent manner. Even more strikingly, these results indicated that some of these genes might be targeted specifically by those miRNAs that demonstrated differential loading onto AGO2 following DNA damage. To provide further evidence for this, we performed a Targetscan analysis$^{140}$ (Figure 26A) and established that many of these genes were in fact predicted targets of p53-activated miRNAs (such as miR-34a, miR-27 and miR-23a/b) and in some cases were targets of miRNAs shown to be more loaded onto AGO2 following DNA damage (such as let-7). For example, RET is a predicted target for miR-34a, CDK6 is a predicted target of miR-34a-c and the let-7 family and FGF18 is predicted target of miR-103a and the mir-200 family (Figure 26A). Furthermore, we performed a seed enrichment analysis to determine whether any AGO2-enriched genes were targeted by p53-regulated miRNAs. Remarkably, the miRNAs to which the AGO2-enriched genes demonstrated the strongest seed-complementary matches were in fact the p53-regulated miRNAs, including miR-34a, let-7, miR-143 and miR-23a which further suggested that these AGO2-enriched genes were potentially targeted by p53-regulated miRNAs (Figure 27).
Figure 26: Many AGO2-enriched genes, including RET, CDK6 and FGF18 are targets of the p53-regulated miRNAs, including in some cases those miRNAs that were more loaded onto AGO2 following DOX treatment and these genes are involved in many cellular signalling pathways that p53 would seek to inhibit after DNA damage.

A, Representation of human RET, CDK6 and FGF18 3’UTRs presenting conserved sites for the p53-regulated miRNA families which are highlighted by a rectangular box (adapted from TargetScan).

B, A pathways analyses determines that many AGO2-enriched genes (highlighted with red boxes) are predicted to play vital roles in p53-regulated signalling pathways such as those involved in cell proliferation, negative regulation of apoptosis, positive regulation of the cell cycle, tumour cell invasion and metastasis, angiogenesis and resistance to chemotherapy.
Finally, when we analysed the signalling networks that the AGO2-enriched genes were involved in, we found that they were vital components in pathways that p53 would seek to inhibit in response to DNA damage, particularly those that promoted tumorigenesis such as proliferation, invasion, metastasis, angiogenesis and resistance to chemotherapy (Figure 26B). Furthermore, a DAVID analysis determined that the group of genes that were more abundant in the RIP-seq samples following DOX treatment were involved in signalling pathways that promoted cell proliferation and cell cycling and supressed apoptosis, whereas the converse was the case for those genes found to be less abundant on AGO2 following DNA damage (Figure 28).

Figure 27: A seed enrichment analysis strongly predicts that AGO2-enriched genes are targets of p53-regulated miRNAs

To evaluate whether the observed remodelling of AGO2-bound mRNAs following induction of DNA damage could have been attributed to an association with modulated miRNAs, we used human target predictions from the Targetscan algorithm\(^1\) and tested each AGO2-enriched (and AGO2-depleted) miRNA for enriched proportion of their predicted targets among AGO2-enriched (and AGO-depleted) transcripts. We limited the database of miRNA target predictions downloaded from the targetscan website (http://www.targetscan.org/) - release 6.1 - to the union set of gene symbols included in our filtered gene datasets (background set, N=11375). Gene redundancy in the prediction database was removed by retaining only the first occurrence of each gene symbol after sorting the database by decreasing 3'UTR length.
In aggregate, these results demonstrated that in its capacity as a guardian against the effects of DNA damage, p53 could fine-tune the activity of a subgroup of genes not through transcriptional means but post-transcriptionally by modulating their targeting by AGO2 through p53-regulated miRNAs. Many genes that were targeted more by the RISC through p53-regulated miRNAs following DNA damage appeared to be crucial components in pathways that promoted cell growth and prevented apoptosis.

3.1.10 An AGO2 PARCLIP analysis illustrates the p53-associated miRNA/mRNA interactome at a genome wide level and establishes the functional effects of miRNA/mRNA binding in terms of p53-regulated signaling networks

The analyses outlined above used RNA-seq and RIP-seq approaches to identify a novel panel of DNA damage-induced miRNAs that were regulated in a p53-dependent and p53–independent manner and the fraction of mRNAs that were regulated at the post-transcriptional level by AGO2 and miRNAs. Furthermore, they enabled a genome wide analysis of the variations in gene expression levels that occurred during the cellular response...
to DNA damage. Additionally, they allowed us to identify a subset of miRNAs and genes that were differentially loaded onto AGO2, and with the aid of Targetscan and seed-enrichment analyses we were able to predict that many of the regulated genes were targeted by the modulated miRNAs. However, there were limitations to this approach which only enabled an indirect or predictive measure of global miRNA/mRNA targeting. Therefore, we wished to provide a more robust analysis of the DNA damage-associated miRNA/mRNA interactome at a genome wide level using an AGO2 PAR-CLIP approach. We hoped that this approach would enable us to confirm which coding genes were regulated by the miRNAs that our RIP-seq analysis had identified as being induced or repressed following DNA damage, and to characterize the sites to which these miRNAs guided the RISC complex to bind to these target transcripts. We also wished to further establish the potential functional effects of any miRNA/mRNA binding on the signaling networks involved in the DNA damage response. Furthermore, we planned to evaluate whether DNA damage modulated the binding of AGO2 on the 3’UTR of mRNAs, independently of the miRNAs modulated by DNA damage. It is feasible that RNA binding proteins modulated by DNA damage might interact with the UTR of specific transcripts and antagonize or aid the interaction of AGO2 with those mRNAs mediating a post-transcriptional regulation. To evaluate this we proceeded to identify whether differential changes in AGO2 loading following DOX treatment was associated with statistically significant differences in the amount of sequence overlapping clusters that constitute the PAR-CLIP signal on discrete regions of the UTRs. This was done in combination with a motif discovery analysis to determine whether particular RNA motifs are associated with these changes in either a p53 dependent or independent manner.

We performed the PAR-CLIP experiment as outlined in the methods section and following deep sequencing we analysed the data using a bioinformatic approach. Briefly, we treated HCT116 p53WT or p53KO cells with DOX or vehicle and performed an AGO-IP after UV-crosslinking in the presence of 4-thiouridine. Following de-crosslinking and RNA extraction, we produced small RNA libraries for each sample and performed a deep sequencing analysis. We produced 3 biological replicates for each experiment and sequenced each in triplicate to produce 9 experimental replicates per sample, in total. In our analysis, we selected for read lengths as short as 13 nucleotides after adapter stripping that contained at least 5 T-to-C mismatches. We defined a cluster as having greater than 5 over-lapping reads and each corresponded to an individual AGO2::mRNA interaction site. We used a kernel density estimate to calculate a ‘signal-to-noise ratio’ in order to determine interaction sites and
matched clusters to the genome to determine the identity of the corresponding gene. The deep sequencing revealed 290,011,684 reads, of which 38% mapped to the genome. We identified 232,964 AGO2-binding sites (clusters) in all PAR-CLIP libraries, 63,385 (27%) of which corresponded to 3’UTRs. These mapped to 5,058 unique genes (4,470 of which were also found in the RNA-Seq) and corresponded with 2,031 human miRNA seeds (nucleotides 2-8). We only considered the AGO2 binding sites that mapped on the 3UTR for further analysis as it has been demonstrated that this interaction co-ordinates post-transcriptional regulation of gene expression but there is conflicting evidence about the possible regulation of gene expression by miRNAs and AGO2 interacting with the coding and 5’UTR regions of the transcripts.

We then performed analyses of selected miRNAs whose seeds matched with each cluster to determine which miRNAs may target these genes. We particularly focused on two subgroups of miRNAs. Firstly, we concentrated on those miRNAs identified in both the small RNA-seq and small RIP-seq analyses as being differentially modulated in a p53-dependent or p53-independent manner in response to DNA damage (illustrated in the venn diagrams in Figure 1E-F) in order to choose the miRNAs regulated most confidently. Secondly, we focused on those miRNAs that were differentially loaded onto AGO2 in a p53-dependent manner following DNA damage. As miRNAs typically target multiple genes within a specific regulatory pathway,400 we used Cytoscape software401 to illustrate the network of genes identified as being high confidence candidates for targeting by each selected miRNA. Cytoscape is a free program developed for the analysis, integration and visualisation of molecular and genetic interaction networks. It processes large amounts of data derived from gene expression profiling, and other functional genomic and proteomic analyses, and integrates them according to an interaction network retrieved for the analysed genes. Data are then visualised in a network graph where biological entities, such as miRNAs, genes or proteins, are represented by nodes, connected with links or edges, which define intermolecular interactions. In our Cytoscape illustrations (Figure 29), miRNAs are represented as triangles that are connected by links to the genes (coloured squares) which our PAR-CLIP analysis identified as being targets of the represented miRNA. Genes shaded in red represent those found to be transcriptionally upregulated in our RNA-seq analysis following DNA damage and those shaded in green depict genes that were transcriptionally down-regulated. Genes shaded in white represent those that showed no differential expression between treatment groups in the RNA-seq analysis but that were identified as targets of the
illustrated miRNA by the PAR-CLIP analysis. It is likely that these genes were insufficiently expressed to be detected in our RNA-seq analysis, rather than them not being actual targets of the illustrated miRNA. Interestingly, this PAR-CLIP analysis showed that in some cases the interaction network for certain miRNAs contained a combination of genes shown to be upregulated and down-regulated in the RNA-seq, whilst in other cases the miRNA in question only targeted genes shown to be down-regulated. For example, let-7 targeted genes such as cyclin D1, polymerase lambda (Poll) and Phorbol-12-Myristate-13-Acetate-Induced Protein 1 (PMAIP1) whose abundance in the RNA-seq analysis was shown to be increased following DOX treatment, but that it also targeted genes such as POGZ, HMGA and PDP2 whose expression decreased. Conversely, miR-3065-5p whose expression was so strongly induced following DNA damage in the RIP-seq was shown only to target genes that were down-regulated in the RNA-seq.

Finally, having identified the gene networks targeted by each selected miRNA, we used DAVID to analyse the cellular signaling pathways in which the genes in each interaction network were involved. Figure 30 illustrates the top GO-term and KEGG analyses of genes identified by PAR-CLIP to be targeted by miRNAs that were upregulated in a p53-dependent manner in the RNA-seq. Interestingly, this analysis demonstrated that amongst these miRNA targets, there was a clear enrichment of genes involved in cell signaling, cell cycling, the DNA damage response and in promoting tumorigenesis. Specifically, our GO-term analysis demonstrated that genes targeted by miR-143-3p (Figure 30C), such as cyclin D1 (CCND1) and the protein regulator of cytokinesis 1 (PRC1), and those targeted by miR-200a-3p (Figure 30D), including Cyclin-Dependent Kinase Inhibitor 1A (CDKN1A or p21), Protein Tyrosine Phosphatase Type IVA Member 1 (PTP4A1 or PRL1), cyclin-dependent kinase 6 (CDK6) and Cell Division Cycle 25B (CDC25B) were important components of such signalling networks. CCND1 encodes a protein belonging to the highly conserved cyclin family that function as regulators of cyclin dependent kinase (CDKs) and whose members are characterized by a dramatic periodicity in protein abundance throughout the cell cycle. Amplification, overexpression or mutation of this gene can alter cell cycle progression and are observed frequently in a variety of malignancies. PRC1 encodes a protein that is involved in cytokinesis and that is highly abundant during the S and G2/M phases of mitosis but not during the G1 phase. It is has been shown to be a substrate of several CDKs and is necessary for polarizing parallel microtubules. CDKN1A regulates cell cycle progression at G1 and is a potent cyclin-dependent kinase inhibitor that binds to and prevents
the activity of cyclin-CDK2 or -CDK4 complexes.\textsuperscript{410,411} It mediates the p53-dependent cell cycle G1 phase arrest in response to stress and plays a regulatory role in S phase DNA replication and DNA damage repair.\textsuperscript{412} CDC25 is a dual specificity phosphatase with three isoforms (A, B and C) that inhibit mitosis under stress conditions through the activation of CDK complexes that regulate cell cycle progression and DNA damage checkpoints.\textsuperscript{413,414}

\textbf{Figure 29: Cytoscape illustrations depicting miRNA:mRNA interaction networks involved in the DNA damage response as identified by a PAR-CLIP analysis}
A) A complete illustration of the miRNA:mRNA interactions activated by DNA damage in p53\textsuperscript{WT} cells. B) A complete illustration of the miRNA:mRNA interactions activated by DNA damage in p53\textsuperscript{null} cells. C) An interaction wheel depicting those genes targeted by let-7f-5p following DNA damage as determined by PAR-CLIP. miRNAs represented by triangles and genes represented by coloured squares (A and B) or circles (C). Red= transcriptionally upregulated by DOX in RNA-seq; Green= transcriptionally down-regulated by DOX in RNA-seq; White= no differential expression between treatment groups in the RNA-seq analysis but that were identified as targets of the illustrated miRNA by the PAR-CLIP analysis. Yellow circle highlights CCND1 as a specific target of let-7f-5p.
Figure 30: Top GO-term and KEGG analyses of genes identified by PAR-CLIP to be targeted by miRNAs that were upregulated in a p53-dependent manner in the RNA-seq
A) GO term analysis showing the function of genes identified by PAR-CLIP to be targeted by any of the miRNAs upregulated in a p53-dependent manner in the RNA-seq. B) KEGG pathway analysis illustrating the cellular signalling pathways that genes targeted by any of the upregulated miRNAs are involved in. C) GO term analysis showing the function of genes identified by PAR-CLIP to be targeted by miR-143-3p. D) GO term analysis showing the function of genes identified by PAR-CLIP to be targeted by miR-200a-3p.
Similarly, genes identified by PAR-CLIP to be targeted by miRNAs that were loaded more onto AGO2, were generally implicated in the promotion of cell division, proliferation and survival and the regulation of cell death in response to DNA damage (Figure 31). Specifically, the let-7 family targeted networks of genes involved in mitosis and cell cycling that included Activin A Receptor Type IB (ACVR1B), CCND1, Type IB Class I Beta-Tubulin (TUBB), Extra Spindle Poles-Like 1 Protein (ESPL1), High Mobility Group AT-Hook 2 (HMGA2), Tumor Suppressor Candidate 2 (TUSC2), Pogo Transposable Element With ZNF Domain (POGZ) and Structural Maintenance Of Chromosomes 1A (SMC1A). Furthermore, let-7 targeted genes involved in the DNA damage response such as Polymerase Lambda (POL), Proliferating Cell Nuclear Antigen (PCNA), RAD18, and SMC1A. miR-24-3p predominantly targeted genes involved in suppressing or promoting apoptosis such as the NUAK Family SNF1-Like Kinase, 2 (NUAK2) and BCL2-Like 11 (BCL2L11), respectively. However we also showed that it targeted genes that repress or stimulate cell proliferation such as Sestrin 1 (SESN1) and epiregulin (EREG), respectively. Interestingly, the miRNAs loaded more onto AGO2 following DNA damage, in particular miR-185-5p and miR-24-3p, appeared to also target mRNAs corresponding to genes that regulate protein phosphorylation, which is relevant for a number of reasons. Firstly, DNA damage pathways engender a large number of phosphorylation events that activate or suppress cell signalling pathways. Secondly, these phosphorylation events also involve p53 and its regulatory partners, including Mdm2 and Mdm4, and such modifications can mediate the induction of p53 and stimulation of its activity.\textsuperscript{415-417} mRNAs targeted by these miRNAs included Protein Kinase Membrane Associated Tyrosine/Threonine 1 (PKMYT1), that encodes a membrane-associated kinase that negatively regulates the cell cycle at the G2/M transition by phosphorylating and inactivating CDK1, and KIT ligand (KITLG) that encodes a protein critical to the regulation of cell survival, proliferation and migration through its ability to promote phosphorylation of PIK3R1, the regulatory subunit of phosphatidylinositol 3-kinase, and subsequent activation of the kinase AKT1.
Figure 31: Top GO-term analyses of genes identified by PAR-CLIP to be targeted by miRNAs that were identified as being loaded more onto AGO2 in the combined RIP-seq/RNA-seq analysis.

A) GO term analysis showing the function of genes identified by PAR-CLIP to be targeted by the let-7 family. B) GO term analysis illustrating the cellular signalling pathways that genes targeted by miR-185-5p. C) GO term analysis showing the function of genes identified by PAR-CLIP to be targeted by miR-24-3p.
Figure 32 illustrates the top GO-term and KEGG analyses of genes identified by PAR-CLIP to be targeted by miRNAs that were upregulated in a p53-independent manner in the RNA-seq. Importantly, these genes are intricately involved in pathways implicated in the DNA damage response in a similar manner to those targeted by miRNAs that were upregulated in a p53-dependent manner, however the target gene sets generally differ. These miRNAs, in particular miR-3065-5p, target mRNAs corresponding to Phosphatase And Tensin Homolog (PTEN), Nucleus Accumbens Associated 1 (NACC1) and PIM1, as well as those mRNAs targeted by p53-dependently upregulated miRNAs such as BBC3 and BCL2L11. NACC1 represses the transcription of the tumour suppressor, Growth Arrest And DNA-Damage-Inducible Gamma Interacting Protein 1 (Gadd45GIP1), and thus promotes cell proliferation and survival as well as oncogenesis.\textsuperscript{418,419} PTEN acts as a dual-specificity protein phosphatase that can function as a tumor suppressor by antagonizing the PI3K-AKT/PKB signalling pathway thereby modulating cell cycle progression and cell survival. Furthermore, PTEN can increase p53 stability and activity through negative regulation of PI3K-AKT/PKB signalling thereby restricting the phosphorylation and movement of Mdm2 into the nucleus.\textsuperscript{420,421} Interestingly, p53 can also directly induce PTEN transcription by interacting with a p53-binding element directly upstream of the PTEN locus.\textsuperscript{204} PIM1 is a serine/threonine kinase that promotes cell proliferation and survival through the positive regulation of MYC transcriptional activity and cell cycle progression, and by phosphorylating and inhibiting pro-apoptotic proteins such as BAD and FOXO3.\textsuperscript{422-426} Furthermore, at physiological levels PIM1 can phosphorylate and activate Mdm2, thereby suppressing p53 activity, but surprisingly, at high levels of expression, PIM1 expression blocks the degradation of both p53 and Mdm2 in a manner that is independent of Mdm2 phosphorylation, leading to increased p53 levels.\textsuperscript{427} Remarkably, the p53-independently upregulated or AGO2 loaded miRNAs, including miR-3065-5p, also targeted mRNA corresponding to the von Hippel-Lindau (VHL) protein. Under normoxic conditions, VHL, through the recruitment of an E3 ubiquitin ligase complex, ubiquitinates hypoxia-inducible factor 1-alpha (HIF-1α) thus inhibiting the transcription of genes involved in angiogenesis, glycolysis, invasion and EMT. However, under stress conditions, including hypoxia, VHL cannot bind HIF-1α which is then free to promote glycolysis and angiogenesis. Interestingly, a recent study in kidney cancer cell lines showed that the VHL protein directly interacted with and positively regulated p53 by inhibiting Mdm2-mediated ubiquitination, and by subsequently recruiting p53-modifying enzymes. Furthermore, VHL-deleted kidney cancer cells showed abnormal cell-cycle arrest or attenuated apoptosis following DNA damage, but
became normal when VHL was restored.\textsuperscript{428,429} This suggested that the VHL protein played a critical role in tumor suppression by participating actively as a component of p53 transactivation complex during DNA damage response. We also found that, following DOX treatment, the p53-independently upregulated miR-4435 targeted a number of mRNAs involved in co-ordinating the DNA damage response including Tripartite Motif Containing 28 (TRIM28) and High Mobility Group AT-Hook 1 (HMGA1). TRIM28 acts predominantly as an inhibitor of transcription but also plays a direct role in suppressing apoptosis. It inhibits E2F1 activity by stimulating the formation of the E2F1-HDAC1 complex and inhibiting E2F1 acetylation and can function to prevent E2F1-mediated apoptosis in the absence of RB1.\textsuperscript{430-433} Furthermore, TRIM28 is an important regulator of CDKN1A/p21(CIP1) and can promote the ubiquitylation and proteosomal degradation of p53/TP53 through a direct interaction with Mdm2.\textsuperscript{434,435} HMGA1 is a non-histone protein involved in many cellular processes, including regulation of inducible gene transcription, integration of retroviruses into chromosomes, and the metastatic progression of cancer cells. However it is also an important regulator of p53 activity that can suppresses p53-mediated mitochondrial-apoptosis through its interaction with Bcl-2,\textsuperscript{436} and inhibit p53 activity by promoting the cytoplasmic re-localisation of the pro-apoptotic activator Homeodomain Interacting Protein Kinase 2 (HIPK2).\textsuperscript{437} Furthermore, HMGA1 can also inhibit the function of the p53 homologues p63 and p73 suggesting the importance of regulating the activity of this protein in the p53-independent response to DNA damage.\textsuperscript{438}

\textbf{Figure 33} illustrates the top GO-term and KEGG analyses of genes identified by PAR-CLIP to be targeted by miRNAs that were down-regulated in a p53-dependent manner in the RNA-seq or that were identified as being loaded less onto AGO2 in the combined RIP-seq/RNA-seq analysis. Again enriched genes were shown to be involved in cellular pathways that regulated cell proliferation and cell division as well as DNA repair and programmed cell death. Interestingly, other pathways including those that coordinated the response to estrogen and steroid hormone stimulation were also amongst the top GO-terms (Figure 9 A). This was not surprising given that many of the proteins involved in these signaling pathways are also important in the cellular response to DNA damage including CCND1, CDKN1A, cyclin A2 (CCNA2) and mitogen-activated protein kinase 1 (MAPK1). Furthermore, although the oestrogen receptor-alpha (ERα) is not expressed in HCT116 cells,\textsuperscript{439} functional ERβ is, and coordinates estrogen-mediated effects on cell growth and proliferation.\textsuperscript{439,440} As one might have expected, mRNAs targeted by miRNAs down-regulated in a p53-dependent manner
following DOX treatment included many of those involved in suppressing the cell cycle and promoting apoptosis or DNA repair such as CDKN1A, CCNA2, SWI/SNF-Related Matrix-Associated Actin-Dependent Regulator Of Chromatin Subfamily A, Containing DEAD/H Box 1 (smarcad1), Eukaryotic Translation Initiation Factor 5A (EIF5A), and Ubiquitin C (UBC), but also anti-apoptotic and pro-proliferative targets such as CCND1, MAPK1, NACC1/2. Remarkably, we also demonstrated that genes targeted by miR-142-5p, such as Dickkopf Homolog 1 (DKK1) were negative regulators of the Wnt-signalling pathway (Figure 33D). Canonical Wnt signaling controls cellular processes such as cell cycle regulation, cell fate determination, EMT, cell motility and metabolism, and pathologic increases in Wnt signalling are frequently implicated in tumorigenesis.441-443 p53 suppresses Wnt activity through a number of mechanisms including the induction of miR-34a that targets positive regulators of the Wnt signalling pathway such as WNT1, WNT3, Low Density Lipoprotein Receptor-Related Protein 6 (LRP6), β-catenin and Lymphoid Enhancer-Binding Factor 1 (LEF1).278,444 Our results demonstrated that miR-142-5p was also involved in the p53-dependent negative regulation of Wnt signalling but that, in this case, p53 mediated its control through a reduction in loading of a miRNA onto AGO2 rather than a transcriptional induction and hence the release of mRNAs involved in suppressing Wnt signalling rather than the targeting of those involved in promoting it.

The GO-term analyses of the mRNAs identified by PAR-CLIP to be targeted by miRNAs that were down-regulated in a p53-independent manner in the RNA-seq are illustrated in Figure 34. These networks generally relate to metabolic and biosynthetic signaling pathways, nucleic acid synthesis and the negative regulation of Wnt signalling that was also demonstrated with miRNAs down-regulated in a p53-dependent manner. Although these pathways are less prominent in the cellular response to double-strand break DNA damage they do have a role particularly in the repair of DNA damage. Interestingly, miR-548d-5p was shown to target chromobox homologue 3 (CBX3) which is a component of heterochromatin that binds to DNA and is recruited to sites of DNA damage and double-strand breaks.445-447
Figure 32: Top GO-term analyses of genes identified by PAR-CLIP to be targeted by miRNAs that were upregulated in a p53-independent manner in the RNA-seq
A) GO term analysis showing the function of genes identified by PAR-CLIP to be targeted by any of the miRNAs upregulated in a p53-independent manner in the RNA-seq. B) GO term analysis showing the function of genes identified by PAR-CLIP to be targeted by miR-3065-3p.
Figure 33: Top GO-term and KEGG analyses of genes identified by PAR-CLIP to be targeted by miRNAs that were downregulated in a p53-dependent manner in the RNA-seq or that were identified as being loaded less onto AGO2 in the combined RIP-seq/RNA-seq analysis

A) GO term analysis showing the function of genes identified by PAR-CLIP to be targeted by any of the miRNAs downregulated in a p53-dependent manner in the RNA-seq. B) KEGG pathway analysis illustrating the cellular signaling pathways that genes targeted by any of the downregulated miRNAs are involved in. C) GO term analysis showing the function of genes identified by PAR-CLIP to be targeted by miR-301a-3p and miR-301b. D) GO term analysis showing the function of genes identified by PAR-CLIP to be targeted by miR-142-5p.
Figure 34: Top GO-term analyses of genes identified by PAR-CLIP to be targeted by miRNAs that were downregulated in a p53-independent manner in the RNA-seq
A) GO term analysis showing the function of genes identified by PAR-CLIP to be targeted by any of the miRNAs downregulated in a p53-independent manner in the RNA-seq. B) GO term analysis showing the function of genes identified by PAR-CLIP to be targeted by miR-548d-5p. D) GO term analysis showing the function of genes identified by PAR-CLIP to be targeted by miR-561-3p.
3.1.11 Luciferase reporter assays confirm that CCND1 and POGZ are targeted by let-7a, let-7d, miR-23a and miR-34a and validate our PARCLIP approach for defining targets of miRNAs regulated upon DNA damage

Next, we performed luciferase reporter assays to validate our PARCLIP approach for defining mRNA targets of miRNAs regulated following DOX treatment. The PARCLIP analysis revealed many genes involved in the DNA damage response that were targeted by miRNAs that were also regulated following DOX treatment. We selected 2 of those genes (CCND1 and POGZ) that are well described as regulators of either cell cycling, mitosis, cell proliferation, DNA repair or programmed cell death and performed 3’UTR luciferase reporter assays (Figure 35). Under conditions of gain-of-function for let-7a, let-7d, miR-23a and miR-34a, we could demonstrate the direct regulation of both of these genes by all 4 tested miRNAs (Figure 35). In fact, overexpression of these miRNAs in HCT116 cells resulted in a significant reduction in the luciferase activity of reporter constructs containing the 3’UTR of either CCND1 or POGZ (Figure 35).

![Luciferase reporter assay graph](image)

**Figure 35: CCND1 and POGZ are direct targets of let-7, miR-23a and miR-34a**
Relative luciferase activity levels were measured 48h after co-transfection of HCT116 cells with CCND1 or POGZ 3’UTR-constructs (pLightSwitch_3UTR GoClone vectors; SwitchGear Genomics, Menlo Park, CA) and the indicated miRNA precursors. Data are mean of three independent experiments ± s.e.m.
3.1.12 p53 regulates the activity of a subgroup of genes involved in the DNA damage response by modulating AGO2 binding on their UTR sequences

We demonstrated previously (section 2.1.9) that p53 post-transcriptionally regulated mRNA expression through differential targeting by AGO2 and that this process was dependent upon p53-regulated miRNAs. Next, we wished to evaluate whether DNA damage modulated the binding of AGO2 on the 3’UTR of mRNAs independently of the miRNAs that were regulated by DNA damage. We believed it was feasible that RNA binding proteins modulated by DNA damage might interact with the 3’UTR of specific transcripts and antagonize or aid their interaction with AGO2 and hence provide an alternative mechanism of post-transcriptional regulation of gene expression.

To evaluate this theory, we initially analysed our RNA-seq data to determine whether DOX treatment induced any significant changes in the expression of known RBPs in a p53-dependent or –independent manner, which may have indicated a role for an RBP in this hypothetical process. Interestingly, we found that DOX treatment induced the differential expression of a number of known RBPs. We confirmed that RBM38 expression significantly increased in a p53-dependent manner following DNA damage (consistent with the literature), as did ZC3H7B abundance. Furthermore, there was a significant reduction in the expression of other RBPs including Heterogenous Nuclear Ribonucleoprotein C (HNRNPC) and Fused in Sarcoma (FUS).

Next, we proceeded to identify whether differential changes in AGO2 loading following DOX treatment was associated with statistically significant differences in the amount of sequence overlapping clusters that constituted the PAR-CLIP signal on discrete regions of the UTRs. This was done in combination with a motif discovery analysis to determine whether particular RNA motifs were associated with these changes in either a p53-dependent or -independent manner and whether any identified motifs corresponded to known RBPs, especially those we had found to be differentially expressed following DNA damage (Figure 36 and Table 15). Remarkably, we found that, in addition to inducing or suppressing AGO2 binding to the 3’UTR of target mRNAs, DOX treatment also induced a p53-dependent remodelling of this interaction. This did not occur exclusively within the miRNA-binding region of the 3’UTR and therefore demonstrated that this differential interaction occurred independently of mRNA targeting by AGO2-bound miRNAs. Figure 37 demonstrates this by
illustrating that DOX treatment altered the amount of sequence overlapping clusters that corresponded to AGO2 binding sites on the 3’UTR of various mRNAs, including CDKN1A, CCDN1, GADD45A GADD45B and SESN1. This occurred to a greater degree in p53WT than p53null HCT116 cell lines. Furthermore, a meme (http://meme.nbcr.net/meme/) de novo motif discovery analysis determined that this binding mechanism was related to a specific RNA motif of 10 nt on the regions of the target 3’UTR that corresponded to the sites that lost or gained AGO2 binding. A particular RNA motif that was highly enriched for GUC at positions 8 to 10 but depleted for G at positions 1 to 7 was specific to p53WT cells (Figure 36). Interestingly, a TOMTOM (http://meme.nbcr.net/meme/cgi-bin/tomtom.cgi) analysis suggested that this RNA motif did not correspond to any known RBP interacting motif (Table 15), although interesting potential candidates include the CELF1 and CELF2 binding proteins, which have been shown to bind to (GUC)n triplets located on mRNA targets that are usually implicated in apoptosis and cell cycle pathways.449,450

These findings demonstrated that p53 post-transcriptionally regulated mRNA expression through modulating the binding of AGO2 on its mRNA targets and that this mechanism was not only mediated by p53-regulated miRNAs bound to AGO2 but also through the reduction, induction and remodelling of AGO2 binding to the 3’UTR of different targets at specific RNA motifs independently from miRNA binding. This provided further insights into the complexity of the AGO2-miRNA-mRNA interaction networks involved in the DNA damage response and their role in modulating gene expression and regulating cellular signaling pathways.
Figure 37: UCSC Genome Browser-adapted density plots
They illustrate that DOX treatment altered the amount of sequence overlapping clusters that corresponded to AGO2 binding sites on the 3'UTR of specific mRNAs including A) CCDN1, B) CDKN1A & C) GADD45A. Pink (DOX treated) and purple (vehicle treated) bars illustrate the location and size of sequence overlapping clusters corresponding to AGO2 binding sites as derived from the PAR-CLIP analysis of treated HCT116 WT cells.
<table>
<thead>
<tr>
<th>Cluster Characteristics and cell type derivation</th>
<th>Enriched Motif</th>
<th>E-value</th>
<th>Number of sites</th>
<th>RBP associated with motif</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clusters located in seed or non-seed regions in 3’UTRs of mRNAs down-regulated in a p53 dependent manner in p53&lt;sup&gt;WT&lt;/sup&gt; cells</td>
<td><img src="motif1.png" alt="Motif" /></td>
<td>1.7e-022</td>
<td>1171</td>
<td>Nil</td>
</tr>
<tr>
<td>Clusters located in seed or non-seed regions in 3’UTRs of mRNAs upregulated in a p53-dependent manner in p53&lt;sup&gt;WT&lt;/sup&gt; cells</td>
<td><img src="motif2.png" alt="Motif" /></td>
<td>2.6e-014</td>
<td>94</td>
<td>Nil</td>
</tr>
<tr>
<td>Clusters located non-seed regions in 3’UTRs of mRNAs down-regulated in a p53 dependent manner in p53&lt;sup&gt;WT&lt;/sup&gt; cells</td>
<td><img src="motif3.png" alt="Motif" /></td>
<td>5.1e-001</td>
<td>212</td>
<td>Nil</td>
</tr>
<tr>
<td>Clusters located in non-seed regions in 3’UTRs of mRNAs upregulated in a p53&lt;sup&gt;WT&lt;/sup&gt;-dependent manner in p53&lt;sup&gt;WT&lt;/sup&gt; cells</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Clusters located in seed regions in 3’UTRs of mRNAs down-regulated in a p53 dependent manner in p53&lt;sup&gt;WT&lt;/sup&gt; cells</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Clusters located in seed regions in 3’UTRs of mRNAs upregulated in a p53&lt;sup&gt;WT&lt;/sup&gt;-dependent manner in p53&lt;sup&gt;WT&lt;/sup&gt; cells</td>
<td><img src="motif4.png" alt="Motif" /></td>
<td>1.3e-008</td>
<td>31</td>
<td>Nil</td>
</tr>
<tr>
<td>Clusters located in seed or non-seed regions in 3’UTRs of mRNAs down-regulated in p53&lt;sup&gt;null&lt;/sup&gt; cells</td>
<td><img src="motif5.png" alt="Motif" /></td>
<td>1.4e-057</td>
<td>59</td>
<td>Nil</td>
</tr>
<tr>
<td>Clusters located in seed or non-seed regions in 3’UTRs of mRNAs upregulated in p53&lt;sup&gt;null&lt;/sup&gt; cells</td>
<td><img src="motif6.png" alt="Motif" /></td>
<td>5.4e-023</td>
<td>69</td>
<td>Nil</td>
</tr>
<tr>
<td>Clusters located in seed or non-seed regions in 3’UTRs of mRNAs down-regulated in p53&lt;sup&gt;null&lt;/sup&gt; cells</td>
<td><img src="motif7.png" alt="Motif" /></td>
<td>3.9e-009</td>
<td>49</td>
<td>Nil</td>
</tr>
<tr>
<td>Clusters located in seed regions in 3’UTRs of mRNAs down-regulated in p53&lt;sup&gt;null&lt;/sup&gt; cells</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Clusters located in seed regions in 3’UTRs of mRNAs upregulated in p53&lt;sup&gt;null&lt;/sup&gt; cells</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Clusters located in non-seed regions in 3’UTRs of mRNAs down-regulated in p53&lt;sup&gt;null&lt;/sup&gt; cells</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Clusters located in non-seed regions in 3’UTRs of mRNAs upregulated in p53&lt;sup&gt;null&lt;/sup&gt; cells</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Clusters located in seed regions in 3’UTRs of mRNAs upregulated in p53&lt;sup&gt;null&lt;/sup&gt; cells</td>
<td><img src="motif8.png" alt="Motif" /></td>
<td>5.0e-018</td>
<td>43</td>
<td>PUM, SNRPA</td>
</tr>
<tr>
<td>Clusters located in seed regions in 3’UTRs of mRNAs upregulated in p53&lt;sup&gt;null&lt;/sup&gt; cells</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

**Table 15: A meme motif discovery analysis and TOMTOM analysis**
Illustrates the 10nt motifs associated with regions of the target 3’UTR (seed or non-seed region) whose binding onto AGO2 increased or decreased following DOX treatment, and describe the RBP associated with each motif. Abbreviations: WT: wild-type; PUM: Pumilio; SNRPA: Small nuclear ribonucleoprotein polypeptide A
3.2 The GAS5-derived snoRNAs are involved in the p53-dependent cellular response to DNA damage

Until very recently there has been little justification for the systematic evaluation of the role of snoRNAs, or snoRNA dysfunction in tumorigenesis or other pathological conditions. This is predominantly due to the fact that in humans, most snoRNAs have been shown to be encoded in the introns of protein-coding and non-protein-coding genes which gave rise to the assumption that these host genes acted solely as cellular housekeepers.\textsuperscript{357,358}

However, recently snoRNAs and their host genes have been implicated in the control of oncogenesis\textsuperscript{359,360} and a number of ‘orphan’ snoRNAs have been identified that have no known rRNA targets.\textsuperscript{361} Furthermore snoRNAs have been located in subcellular compartments other than the nucleolus,\textsuperscript{361} which supports the concept that they may regulate other molecules and have additional cellular functions other than as regulators of the ribosome.\textsuperscript{362} Even more strikingly, other studies have reported that mature snoRNAs may undergo further cellular processing to form smaller sdRNAs with miRNA-like functions.\textsuperscript{347,363-366} Additionally, snoRNA expression has been shown to be as variable as miRNA expression in human tumour samples and normalising miRNA PCR expression data to these snoRNAs introduced bias in associations between miRNAs and outcome.\textsuperscript{367}

The potential role of snoRNAs in tumorigenesis was first suggested by Chang \textit{et al.} (2002) who demonstrated substantially lower snoRNA expression in meningiomas compared with normal brains.\textsuperscript{376} Subsequent studies have linked dysregulation in the expression of various snoRNAs to the development of a number of malignancies including non-small-cell lung cancer,\textsuperscript{373,374} prostate cancer\textsuperscript{375} and breast cancer.\textsuperscript{360} Interestingly, the \textit{GAS5} gene that hosts ten intronic snoRNAs and a lncRNA, has been implicated in oncogenesis and in the regulation of cell survival by inducing or sensitizing cells to apoptosis.\textsuperscript{379-381} \textit{GAS5} mRNA levels were significantly decreased in breast cancer samples compared with adjacent unaffected normal breast epithelial tissues which suggests a role for \textit{GAS5} as a tumour-suppressor gene. However, little is known about whether it is the \textit{GAS5}-derived snoRNAs or lncRNA that are responsible for these effects, although a previous study has suggested that three of the \textit{GAS5}-derived snoRNAs (U44, U74 and U78) may also give rise to sdRNAs with miRNA-like function\textsuperscript{347} although no mechanism or specific targets were shown and there was no consistency across cell lines. In work related to results described earlier in this chapter, we discovered that expression levels of the \textit{GAS5}-derived snoRNAs U44 and U47 increased in...
colon cancer cell lines following doxorubicin treatment and that this occurred in a p53-dependent manner. Given the previously described role of GAS5 in the regulation of apoptosis and the well documented role for p53 in the same process, we aimed to further investigate the relationship between p53 and the GAS5 snoRNAs to gain further insight into their potential role in cell survival and oncogenesis in colorectal cancer both in vivo and in vitro.

3.2.1 Doxorubicin-induced DNA damage increases GAS5-derived snoRNA expression in a p53 dependent manner in colorectal cancer cell lines

We treated HCT116 p53WT and HCT116 p53KO cells with doxorubicin in order to induce DNA damage, and used RT-qPCR to measure the changes induced in the expression levels of various small RNAs, particularly the GAS5-derived snoRNAs U44 and U47. Doxorubicin treatment in HCT116 p53WT cell lines led to a significant induction in the expression of the GAS5-derived snoRNAs U44 (P < 0.01) and U47 (P < 0.01) when compared to treatment with a control vehicle, but there was no significant change in levels of the non-GAS5-associated snoRNA U19 (Figure 38) or the snRNA U6 (data not shown). Doxorubicin treatment did not significantly increase GAS5-derived snoRNA expression in HCT116 p53KO cells (Figure 9), suggesting that DNA damage induced the expression of the GAS5-derived snoRNAs in a p53-dependent manner. Doxorubicin treatment of HCT116 p53WT cells also significantly increased the expression of miR-34a (P ≤ 0.006), used as a positive control, although the size of the fold change varied depending on which small RNA was selected to normalize expression levels to (Figure 38 & 39). Following 24 hours of doxorubicin treatment, miR-34a expression increased significantly by 2.8-fold and 2.6-fold (P ≤ 0.006 for both) when levels were normalised to U6 and U19 respectively (Figure 10). However, although still statistically significant, the fold-changes in miR-34a expression levels were much smaller (1.9-fold; P < 0.01) when the GAS5-derived snoRNAs U44 and U47 were used for normalisation (Figure 39). Similar differences were seen when p21 was used as a positive control (data not shown).
Figure 38: DOX induces the expression of the G455-derived snoRNAs
Relative levels of (A) miR-34a, (B) U19 snoRNA, (C) U44 snoRNA and (D) U47 snoRNA were measured by RT-qPCR in p53^{WT} HCT116 cell lines and p53^{KO} HCT116 cell lines treated with either doxorubicin (at a final concentration 0.2μg/ml) or vehicle for 24 hours. Levels were normalised to U6 snRNA levels and data are presented relative to the vehicle treated cells ± s.e.m (each of them performed in triplicate; Student’s t test: *P < 0.01, **P ≤ 0.006).
Figure 39: The G455-derived snoRNAs are not suitable housekeeping genes for use in experiments where DOX treatment
Relative levels of miR-34a normalised to (A) U6 snRNA, (B) U47 snoRNA, (C) U44 snoRNA and (D) U19 snoRNA were measured by RT-qPCR in p53^WT HCT116 cell lines and p53^KO HCT116 cell lines treated with either doxorubicin (at a final concentration 0.2ug/ml) or vehicle for 24 hours. Data are presented relative to the vehicle treated cells ± s.e.m (each of them performed in triplicate; Student’s t test: *P < 0.01, **P ≤ 0.006).
3.2.2 GAS5-derived snoRNA expression varies between normal and malignant colorectal fresh non-microdissected tissue in a p53-dependent manner

Having identified a relationship between p53 activity and the expression of GAS5-snoRNAs in colorectal cancer cell lines, we wished to investigate whether a similar association existed in human colorectal tissue, and in particular in colorectal tumour samples. We collected paired samples of fresh frozen colorectal tumour tissue and corresponding normal colorectal tissue from 20 individual patients and used RT-qPCR to measure miR-34a and snoRNA expression levels in these samples. We found significant differences in the expression levels of the GAS5-derived snoRNAs between paired samples of fresh frozen normal colorectal tissue and colorectal tumour from the same patient (P< 0.01; Figure 40). snoRNA levels were significantly higher in tumours compared to the corresponding normal colorectal tissue in 85% of patient samples, but were significantly lower in 15%. There was no significant difference in snRNA U6 levels between paired normal and tumour samples. Interestingly, miR-34a levels were also significantly higher in patient tumour samples when compared to their corresponding normal colorectal tissue samples (P≤ 0.0006; Figure 40). We then measured p53 expression levels in the paired normal colorectal tissue and colorectal tumour samples using Western blotting (Figure 41A-C). p53 levels were significantly higher in colorectal tumours compared to their corresponding normal colorectal tissue samples (P< 0.01; Figure 41D).

Using the same samples, we calculated Pearson’s correlation coefficients, comparing p53 expression levels with snoRNA U44 and U47 levels, to determine if any relationship existed between GAS5-derived snoRNA levels and p53 in vivo in humans. We found a strong positive correlation between p53 expression levels and the levels of both snoRNA U44 (Pearson Correlation= 0.64; R^2 linear = 0.41) and snoRNA U47 (Pearson Correlation= 0.69; R^2 linear = 0.49) in colorectal tumour samples (Figure 42A). We also calculated Pearson’s correlation coefficients to compare miR-34a expression levels with snoRNA U44 and U47 levels, to determine if any relationship existed between the levels of GAS5-derived snoRNAs and p53-regulated miRNAs in humans. This was also performed to provide evidence in support of the use of miR-34a as a surrogate marker for p53 in this context for additional experiments using RNA derived from microdissected FFPE tissue samples in which p53 levels were not measurable by Western blotting. Interestingly, we found a strong positive
correlation between miR-34a expression levels and the levels of both snoRNA U44 (Pearson Correlation= 0.73; $R^2$ linear = 0.53) and snoRNA U47 (Pearson Correlation= 0.66; $R^2$ linear = 0.43) in colorectal tumour samples (Figure 42B).

Figure 40: Expression of the G4S5-derived snoRNAs is greater in colorectal tumours than in benign colon tissue
A Box plot comparing the relative expression levels of miR-34a, U44 snoRNA and U47 snoRNA between paired colorectal tumour (T) and normal colorectal (N) fresh frozen tissue samples. (Student’s t test *P < 0.01, ***P ≤ 0.0006)
Figure 41: A comparison of p53 expression levels between paired normal colorectal and colorectal tumour tissue samples

(A & B) Western blot’s showing p53 levels in the first 10 normal colorectal (A) and colorectal tumour (B) tissue samples. GAPDH was used as a loading control. (C) Column chart demonstrating the fold changes in p53 expression levels shown in the Western blots (A & B) normalised to GAPDH and calculated using ImageJ software. (D) A box plot comparing the relative expression levels of p53 between all 25 paired colorectal tumour (T) and normal colorectal (N) tissue samples (Student’s t test *P < 0.01).
Figure 42: A positive correlation exists between GAS5-derived snoRNA levels and p53 expression in colorectal tissue samples

Graphs showing Pearson correlation analyses of the relationship between (A) p53 levels or (B) miR-34a levels and the snoRNAs U44 and U47 in colorectal tumour tissue samples (A & B).
3.2.3 \textit{GAS5}-derived snoRNA expression varies between normal, pre-malignant and malignant microdissected FFPE colorectal tissue and levels correlate with miR-34a expression

There is much debate as to the accuracy of RNA and gene expression studies that use non-microdissected tumour samples, due to the possible effects that the cellular components of the surrounding stroma can have on the levels of the measured molecule. We therefore aimed to perform further experiments in microdissected tissue samples to support the results above. We collected 60 unpaired FFPE colorectal tissue samples consisting of 20 normal mucosa, 20 adenoma and 20 tumour specimens. We microdissected the required portions after H&E staining, performed RNA extraction and measured small RNA expression levels by RT-qPCR. We found significantly higher levels of \textit{miR-34a} (P ≤ 0.006), snoRNA U44 (P ≤ 0.0005) and snoRNA U47 (P ≤ 0.0005) in adenoma samples when compared with normal mucosa samples (Figure 43A). Furthermore, the expression of all 3 small RNAs was significantly higher in tumour samples compared to adenoma or normal mucosa samples (Figure 14A). In addition, p53 levels measured by immunohistochemistry and given as a p53 score of 0-3, were higher in tumour samples (80\%= score of 3, 20\%= score of 2) and adenoma samples (50\%= score of 3, 30\% score of 2, 20\% score of 1) than normal tissue (100\%= score of 0).

Using the same samples, we then calculated Pearson’s correlation coefficients to compare \textit{miR-34a} expression levels and snoRNA U44 and U47 levels to determine if the relationship demonstrated in non-microdissected samples between \textit{GAS5}-derived snoRNA levels and p53 (\textit{miR-34a} being used here as a surrogate marker for p53) was also seen in microdissected colorectal tumours. We found a strong positive correlation between \textit{miR-34a} expression levels and the levels of both snoRNA U44 (Pearson Correlation= 0.69; \( R^2 \) linear = 0.47) and snoRNA U47 (Pearson Correlation= 0.67; \( R^2 \) linear = 0.45) in colorectal tumour samples (Figure 43B).
Figure 43: In microdissected colon samples the GAS5-derived snoRNAs are expressed more in malignant and pre-malignant tissue than benign tissue and levels correlate with p53 expression.

A, RT-qPCR was used to measure the relative expression levels of miR-34a, snoRNA U44 and snoRNA U47 in microdissected human tissue samples corresponding to normal colorectal tissue (N), colorectal adenoma (A) and colorectal tumours (T) (Student’s t test *P < 0.05,**P ≤ 0.006,***P ≤ 0.0005). B, A Pearson’s correlation analysis was performed to determine the relationship between miR-34a levels and the snoRNAs U44 and U47 in microdissected colorectal FFPE tumour tissue samples.
3.2.4 The expression of GAS5-derived snoRNAs is not affected in colorectal cancer cell lines in which DICER has been knocked-down and therefore do not appear to be processed by DICER

In view of the findings described in previous studies which demonstrated that snoRNAs can be converted to sdRNAs with miRNA-like functions and that processing by DICER is the most likely mechanism responsible, we wished to assess the effect of DICER knock-down on the expression of GAS5-derived snoRNAs following DNA damage. To achieve this we used RT-qPCR to compare changes in the expression of the snoRNAs U44 and U47 following doxorubicin treatment in the colorectal cancer cell lines DLD1 and RKO in their wild-type (WT) form and in a form in which DICER had been stably knocked down (KD). Interestingly, we found that although DICER knock-down led to a statistically significant reduction in miR-34a levels (P ≤ 0.006) in both DLD1 and RKO cell lines (as one would expect), there was no effect on the levels of snoRNA U44 or snoRNA U47 (Figure 44). This suggests that the function of the GAS5-derived snoRNAs in the p53-regulated response to DNA damage does not involve their conversion into sdRNAs with miRNA-like function, or that if it does, DICER is not involved in this processing step.

In aggregate, these findings demonstrate that a relationship exists between p53 activity and the expression of the GAS5-derived snoRNAs in colorectal cancer cell lines and human colorectal tissue. Furthermore, they suggest that transcription of the GAS5 gene is directly regulated by p53, although chromatin immunoprecipitation (ChIP) studies using DOX induced DNA damage would be required to confirm this. Interestingly, analysis of previous p53 chromatin immunoprecipitation sequencing (ChIP-seq) experiments performed in HCT116 cells indicates the presence of a significant peak of p53 interaction in two independent experiments involving p53 activation induced by Nutlin3 or 5'fluorouracil (5FU), at the same position, approximately 800bp away from the GAS5 transcriptional start site (TSS), indicating that p53 does directly control GAS5 transcription. Although no functional studies have been performed, these findings suggest an important role for the GAS5-derived snoRNAs in the p53-regulated cellular response to DNA damage in colorectal cancer cell lines and in p53-associated signalling pathways in human colorectal tissue and colorectal cancer. It appears that their role is not related to a miRNA-like function, as in our studies, they did not appear to be processed by DICER into sdRNAs. Furthermore, our results
demonstrate that the GAS5-derived snoRNAs are not appropriate for use as ‘housekeeping genes’ for normalisation of RT-qPCR measurements of small RNAs in experiments involving DNA damage because fold changes will be lower than their true relative value which would lead to inaccurate results. Our data suggest that the snRNA U6 or snoRNA U19 would be more appropriate for normalisation in such experiments.

Figure 44: The GAS5-derived snoRNAs are not processed by DICER
Relative levels of U44 snoRNA (red), U47 snoRNA (blue) and miR-34a (purple) were measured by RT-qPCR in DLD1 DICER<sup>WT</sup> cell lines, DLD1 DICER<sup>KD</sup> cell lines, RKO DICER<sup>WT</sup> cell lines, RKO DICER<sup>KD</sup> cell lines treated with either doxorubicin (at a final concentration 0.2ug/ml) or vehicle for 24 hours. Data are presented relative to the vehicle treated corresponding cell lines (dotted line) ± s.e.m (each of them performed in triplicate; Student’s t test: *P < 0.01, **P ≤ 0.006).
3.3 Optimisation of experimental techniques

3.3.1 PAR-CLIP protocol modifications

The original PAR-CLIP protocol published by Hafner et al. (2010) involved the use of two RNA digestion steps to reduce the size of the RBP-bound RNA and thus to enable sequencing of the bound fraction and determination of its identity and binding sequence. Firstly, the cell lysates underwent an RNA digestion step using RNase T1 (Fermentas 1,000U/u) at a final concentration of 1U/ul, followed by incubation at 22°C for 15 minutes. The second RNA digestion step which followed the IP stage of the protocol used RNase T1 (Fermentas 1,000U/u) at a final concentration of 100U/ul followed by incubation at 22°C for 15 minutes. However, although these concentrations allowed the digestion of RBP-bound fragments of RNA down to the appropriate size for the cell types and RBP used in the Hafner paper, we found that this caused overdigestion of the RNA bound to AGO2 in our PAR-CLIP experiment so that the resulting size of the RNA fragments at the end of the small RNA library preparation was too small. More specifically, in our study the same RNase concentrations used in the original Hafner study produced RNA fragments of only 10-13 bp which when converted to cDNA using the Illumina Truseq small RNA library preparation kit could not be used for sequencing as they could not be accurately mapped to the genome and hence their identity could not be established with any certainty. This is illustrated by the bioanalyser traces in Figure 45. The final construct recovery step of the Illumina Truseq small RNA library protocol should produce cDNA that are 143-160 nucleotides (nt) in length which consist of 19-22 nt corresponding to the AGO2-bound miRNA/digested mRNA and 125 nt corresponding to the adaptors bound to either side of the RNA during the ‘ligation of adaptors’ step in the library protocol that preceded the reverse transcription step. However, as the over-digestion of the RNA during the PAR-CLIP step led to the production of RNA fragments that were only 10-13 nt rather than 19-22 nt in length then the resulting fragments of cDNA produced during the subsequent library preparation steps were around 130 nt in length (Figure 45A) and could not therefore be sequenced effectively. We therefore repeated the PAR-CLIP protocol on a number of occasions, modifying the concentrations of RNase used in each of the RNA digestion steps, as well as the length of incubation times. Furthermore, in case the over-digestion was the result of RNase activity due to unintentional contamination of the RNA samples with RNase during the various steps of the procedure, we
were also increasingly vigilant to ensure all apparatus was decontaminated with RNase zap before use. After many repetitions of the PAR-CLIP and library protocols we found that the optimal concentrations of RNase to be used in our PAR-CLIP experiment using HCT116 cells and immunoprecipitating for AGO2 were 1U/ul in the first digestion step and 10U/ul (rather than 100U/ul) in the second digestion step. This enabled the production of RNA fragments that were predominantly 19-30 nucleotides in length, that could then be used successfully to produce a small RNA library that was enriched for cDNA that was 143-160 nt in length and that could therefore be successfully sequenced and then after adaptor stripping could allow the resulting fragments to be mapped to the genome and identified (Figure 45B).

3.3.2 Modifications to the small RNA and mRNA library preparation protocols for AGO2 IP and PAR-CLIP derived RNA

The Illumina Truseq small RNA library preparation protocol states that the starting quantity of RNA should be 1ug diluted in 5ul of RNase free water. The Illumina Truseq mRNA library preparation states that the starting quantity of RNA should be 1-4ug diluted in 50ul of RNase free water. However there was no guidance in either protocol on the amount of RNA that was required as a starting concentration for RNA derived from an AGO2 IP or PAR-CLIP experiment. Therefore we initially performed a number of library preparations starting with increasing concentrations of RNA derived from both the AGO2 IP and PAR-CLIP protocols and assessed the resulting yield of cDNA using a bioanalyser (Figure 46). We found that for the small RNA libraries and mRNA libraries produced using RNA from the AGO2 IP, the minimum amount of RNA required as a starting quantity was 250ng. Lower concentrations of RNA did not allow an effective reverse transcription step and therefore there was no detectable cDNA or insufficient amounts of cDNA at the end of the library to allow efficient sequencing to be performed. This is because the RNA produced from AGO2 IP experiments is more enriched for miRNAs and mRNAs and more deplete in tRNA and rRNA than total RNA or RNA derived from the input samples of an IP experiment. Therefore, for the cDNA library preparation for RNA derived from our experiments we used 4ug of starting material for the total RNA or input samples and 250ng of RNA for samples derived from the AGO2 IP experiments. For the small RNA seq library this meant that we required a final concentration of RNA from the AGO2 IP of at least 50ng/ul to enable a
quantity of starting RNA of 250ng as the volume required in the first step of the small RNA library is 5ul. The quantities of reagents, including the number of cells used and the quantities of sepharose beads and AGO2/IgG antibodies, described in the methods section of this thesis that describes the AGO2 IP protocol are those that we found were required to most consistently yield a concentration of at least 50ng/ul of RNA that is required for the small RNA library preparation step.

The PAR-CLIP experiment created another conundrum with regards to the starting amount of RNA required for the subsequent small RNA library preparation step prior to next generation sequencing. The PAR-CLIP experiment produces a final RNA product that is heavily degraded due to the RNase digestion steps used during the protocol. The concentration of RNA in each sample is quantified using a nanodrop measurement at the end of the PAR-CLIP experiment prior to the library preparation. However, we found that the concentration of starting RNA required for the small RNA library to allow an effective reverse transcription and hence to yield sufficient cDNA to allow effective sequencing was far higher for the PAR-CLIP experiment than for the AGO2 IP experiment. In order to ensure the RNA sample at the end of the PAR-CLIP experiment was as concentrated as possible, we dissolved the final RNA pellet in only 6ul of RNase free water. This allowed 1ul to be used for a nanodrop measurement to ensure RNA was present and of good quality and the remaining 5ul to be used for the small RNA library preparation that requires a 5ul starting volume. The starting number of cells used in our PAR-CLIP experiments as well as the amount of dynabeads and antibody used were calculated through a number of experiments performed to determine exactly what was required to enable a sufficient yield for effective library preparation and sequencing.

We also had to optimise the Truseq mRNA protocol by performing an additional modification step after the final PCR stage. Despite very careful use of the ‘clean up beads’ during the mRNA library preparation that aims to ensure that no significant adaptor contamination was present in the final cDNA sample, we had difficulties in ensuring that all the samples were free of adaptor contamination. The adaptors are used during the library preparation but if free adaptors (unbound to RNA) are left in the sample at the PCR stage then these are also amplified along with the cDNA. This means that a significant amount of DNA within the samples will have derived from the adaptors and not from reverse transcription and PCR of the RNA produced in the AGO2 IP that we wished to measure by Next-generation sequencing. If adaptor dimers are sequenced with the cDNA derived from
the RNA we wish to measure then they absorb a large number of reads during the sequencing process which leads to inaccurate and often uninterpretable results. Therefore it is imperative to keep the concentration of adaptor dimers to below 10% of the concentration of cDNA derived from the RNA produced in the AGO2 IP. The levels of adaptors can be seen by bioanalyser analysis and compared with the levels of cDNA derived from mRNA during the same process. In order to ensure that the adaptor concentration was as low as possible, we performed an additional gel cutting procedure followed by a cDNA gel extraction step. We ran the samples of RNA derived from the AGO2 IP on a 150ml 1% agarose gel and then cut the band at 200-500bp that corresponded to the mRNA but ensured that we did not cut any bands at 120-130bp that corresponded to adaptor dimers. This ensured that there was no adaptor contamination in our samples prior to sequencing and is illustrated in Figure 47. Interestingly, but perhaps unsurprisingly, we found that adaptor contamination was a particular issue in the AGO2 IP samples and not the input or total RNA samples. This can be explained by the fact that the AGO2 IP samples contained significantly less quantities of starting RNA that the input or total RNA samples but the same amount of adaptors were used in all the library preparations. Therefore in the AGO2 IP samples, there was less material to ‘mop up’ the unbound adaptors during the library preparation and so more adaptor dimers formed.
Figure 45: Bionalyser traces of cDNA derived from RNA produced in the PAR-CLIP experiment that was then used to create a cDNA library of small RNAs. (A) cDNA library produced from a PAR-CLIP experiment using RNase at a concentration of 1U/ul followed by 100U/ul that demonstrates an average bp peak of 132 bp. (B) cDNA library produced from a PAR-CLIP experiment using RNase at a concentration of 1U/ul followed by 10U/ul that demonstrates an average bp peak of 141-158bp. The peaks at 35 and 10380 nt (bp) correspond to the marker peaks used in the bioanalyser protocol.
Figure 46: Bionalyser traces of cDNA derived from RNA produced in the AGO2 IP experiment that was then used to create a cDNA library of small RNAs. (A) cDNA library produced from an AGO2 experiment that used 100ng as a starting quantity of RNA and did not therefore produce a peak of cDNA at 143-160 bp demonstrating that no cDNA corresponding to miRNAs (19-22nt) had been produced in the library. (B) cDNA library produced from an AGO2 experiment that used 250ng as a starting quantity of RNA and did produce a peak of cDNA at 143-160 bp demonstrating that cDNA corresponding to miRNAs (19-22nt) had been produced in the library. The peaks at 35 and 10380 nt (bp) correspond to the marker peaks used in the bioanalyser protocol. The peaks at 101 and 123 bp correspond to adaptor dimers.
Figure 47: Bionalyser traces of cDNA derived from RNA produced in the AGO2 IP experiment that was then used to create a cDNA library of mRNAs. (A) cDNA library produced from an AGO2 experiment following which no gel clean up procedure was used prior to performing a bioanalyser measurement of the yielded products. (B) cDNA library produced from an AGO2 experiment following which a gel clean up procedure was used prior to performing a bioanalyser measurement of the yielded products. The peaks at 128 bp correspond to the adaptor dimers and are not present in the sample after gel cutting (B). The peak between 200 and 500 bp correspond to the mRNA product.
Chapter 4: Discussion
4.1 The p53-regulated small RNA network and its role in the DNA damage response

4.1.1. We identified a novel group of miRNAs regulated by DNA damage and found that this stress stimulus induced the differential loading of specific miRNAs onto AGO2, although the mechanism responsible requires further evaluation

We selected an RNA-seq approach rather than a microarray approach in order to identify novel DNA-damage induced or supressed miRNAs. Although informative, a microarray approach would not have allowed the most detailed analysis at the transcriptome wide level because such platforms suffer from a fundamental ‘design bias’ in that they can only return results from those regions for which probes have been constructed and consequently are only as good as the databases from which they are devised. The list of known miRNAs on the database, miRBase, is continually updated and therefore even the most up-to-date array platform would not have enabled the detection of every known miRNA. Conversely, RNA-seq provided a tool by which the whole transcriptome could be studied. We confirmed the p53-dependent up-regulation of miR-34a, miR-143, miR-200 and miR-107, consistent with previous findings and which indicated the reliability of our methods, however we were not able to confirm a significant fold-change for the previously described p53-regulated miRNAs miR-15a and miR-16. This may be the result of different cell lines used in our experiments than in some previously described studies or the use of a different DNA damaging agent but most likely the result of certain minor biases in our analytical techniques. However, on the whole our results were very consistent with published data and were validated by RT-qPCR and are therefore reliable in our view. Our approach enabled us to identify a group of miRNAs not previously demonstrated to be regulated by DNA-damage including the p53-dependently up-regulated miR-135b-3p, miR-4485 and a host of down-regulated miRNAs including miR-4417, miR-548d-3p and miR-1266-5p. Prior to our study, there was a paucity of data surrounding those miRNAs that were down-regulated in a p53-dependent manner following DNA damage, and strikingly many were found to be negative regulators of the DNA damage response. Interestingly, our RNA-Seq approach permitted us to identify for the first time, miR-3065-3p and miR-3065-5p as the miRNAs most significantly up-regulated by DNA damage, and this occurred in a p53
independent manner, indicating that it could be very important in the DNA damage response. It is possible that the p53 family members, p63 and p73 are involved in the transcriptional regulation of miRNAs such as miR-3065-3p that are differentially expressed following DNA damage in a p53-independent manner. By combining these results with our PAR-CLIP data we were then able to illustrate in more detail the role of these identified miRNAs in the regulation of gene networks responsible for mediating the DNA damage response.

In addition, we demonstrated that DNA damage induced the differential loading of certain miRNAs, such as the let-7 family members, onto AGO2 and that this occurred in a p53-dependent manner. Furthermore, it was only demonstrated in p53 wild-type cell lines (HCT116 p53WT and RKO) but not in DLD1 cells that are known to express one allele that encodes a mutant form of p53 (a missense C > T mutation at position 241 that results in a Ser > Phe substitution) and one p53 allele that is silenced. Due to time constraints, we were not able to perform further experiments to determine the mechanism involved here and this is discussed in the ‘future work’ chapter of this thesis, however we were able to hypothesise on potential mechanisms responsible for these findings. Initially, we postulated whether this phenomenon may have been due to an increase in AGO2 protein abundance following DNA damage that might have occurred in a p53-dependent manner. Potential mechanisms responsible included a p53-mediated increase in transcription of the AGO2 gene, EIF2C2, or a p53-regulated decrease in degradation of AGO2 through autophagy. Previous studies have demonstrated that both DICER and AGO2 were selected for autophagy by the selective autophagy receptor, NDP52, and that this could affect miRNA abundance and function.452,453 Although autophagy has been shown to be a process regulated by p53 in response to DNA damage, we could not find any evidence in the literature that demonstrated a p53-dependent reduction in autophagy regulation of AGO2 following DNA damage. Furthermore, we could find no evidence of a p53-dependent regulation of AGO2 transcription either in the literature or our RNA-seq analysis. Crucially, our Western blot analysis demonstrated that DOX did not induce a change in cellular AGO2 expression at the protein level suggesting the AGO2 loading affect was not related to a change in AGO2 abundance. Furthermore, it would have been very unlikely that an increase in AGO2 expression would have resulted in a selective rather than global increase in miRNA loading. A more probable explanation was that there was a p53-dependent change in AGO2 loading of certain miRNAs due to either a specific conformational change in AGO2 or the specific binding of these miRNAs to another
molecule. We postulated that RNA-binding proteins (RBP) were the most likely candidates here.

RBPs have been shown to have a role in modulating miRNA processing and function.\textsuperscript{75,454-456} The heterogeneous nuclear ribonucleoprotein A1 (hnRNP A1) was the first miRNA-binding protein to be identified.\textsuperscript{454} Previous studies demonstrated hnRNP A1 was involved in the packaging, splicing and nucleocytoplasmic transport of pre-mRNA, but further work showed that it could also bind to the stem loop of pri-miR-18a and facilitate its processing by DROSHA to pre-miR-18a.\textsuperscript{454} LIN-28 has been shown to impair let-7 processing at both the DROSHA and DICER levels by interacting with the pri-let-7 stem loop\textsuperscript{75} and by inducing 3'-terminal uridylation of pre-let-7,\textsuperscript{455} respectively. The KH-type splicing regulatory protein (KSRP) binds with high affinity via its KH3 domain to G-rich stretches in the terminal loop of many pri- and pre-miRNAs (e.g. let-7a-d, miR-21, miR-16) and also binds to and serves as a component of DROSHA and DICER complexes, thereby increasing DROSHA and DICER processing by optimising the positioning of miRNA constructs with these RNAse III enzymes.\textsuperscript{456} We did not find a significant change in the abundance of any of the mRNAs corresponding to these proteins in our RNA-seq analysis making a role for them in the AGO2 loading affect less likely. Conversely, it is possible that such RBPs may in fact have been involved here but that they were activated by p53-induced post-translational modifications rather than a p53-dependent change in their abundance, as the p53 signaling pathway has been shown to induce many post-transcriptional changes on proteins. Interestingly, stress stimuli have been shown to alter the phosphorylation status of hnRNP K and SUMO modification of hnRNP C/M led to conformational changes.\textsuperscript{457} Strikingly, a recent study used a PARCLIP analysis followed by fluorescence anisotropy-based binding assays to demonstrate that HUR interacted with let-7 and functionally affected the association between AGO2 and let-7 loading.\textsuperscript{458} To our knowledge, this study and our data were the first to demonstrate a mechanism whereby the loading of miRNAs onto AGO2 can be modulated and we were the first to demonstrate this within the context of the p53-mediated response to DNA damage. Whether HUR is involved in this process demonstrated herein remains to be seen.

Subsequently, we reviewed whether p53 itself could be a direct mediator of the increased loading of miRNAs onto AGO2 in response to DNA damage. Although historically, p53 has been considered to be a sequence-specific DNA-binding transcription factor, reports over
recent years have described protein-protein and protein-RNA interactions involving p53 as well. p53 has been shown to regulate miRNA processing through protein-protein interactions at the DROSHA level through binding with DDX5 and the p53 family members, p63 and p73, also appear to regulate components of miRNA processing (i.e. Drosha-DGCR8, Dicer-TRBP2, and Argonaute proteins). p53–RNA interactions have been shown to be mediated by a p53 nucleic acid-binding domain independent of the sequence-specific core domain responsible for DNA recognition. For example, p53 interacted with RNA via sequence-specific and sequence-nonspecific mRNA binding, covalent complexes and sequence-nonspecific RNA annealing, however, some have argued whether interactions between p53 and RNA are in fact physiologically relevant. We felt it was unlikely that p53 bound directly to the let-7 miRNAs to facilitate their loading onto AGO2, but rather that p53 interacted with a protein that enabled this process. Furthermore, we postulated that p53 might interact with AGO2 itself and induce a conformational change to facilitate the loading of specific miRNAs. In support of this hypothesis, a recent study in Drosophila used a genome-scale in vitro expression cloning approach to identify previously unreported p53 interactors, and found 91 candidates including both AGO1 and AGO2. Although this interaction between p53 and AGO2 was demonstrated in Drosophila and not human cell lines, the authors went on to analyse for evolutionary conservation of the p53-interacting proteins and tested 41 mammalian orthologs and found that 37 bound to one or more p53-family members when overexpressed in human cells. These data support a possible direct role for p53 in the AGO2 loading effect that we demonstrated and suggest that following DNA damage p53 bound directly to AGO2 and induced a conformational change that encouraged the loading of certain miRNAs such as the let-7 family members. Another possibility was that p53 bound indirectly to AGO2 through an interaction with a protein able to bind to the RISC complex. A potential candidate protein is the Activator of PKR (PACT), a 250 kDa Ring-finger containing DICER-associated RBP that along with TRBP was shown to be a key regulatory factor involved in substrate and cleavage specificity during miRNA and siRNA production. Interestingly, recombinant forms of PACT bound to wild-type p53, interfered with its binding to DNA and acted as a negative regulator, but was unable to interact with mutant forms of p53. It is possible that in our experiments, DNA damage activated wild-type p53 to interact with the RISC complex through its ability to bind to DICER-associated PACT and induced a conformational change that modulated the loading of certain miRNAs onto the RISC. In keeping with this theory, our finding that the mutated form of p53 in DLD1 cells was unable to induce an increase in AGO2 loading might be
explained by its inability to bind to PACT on the RISC complex. Interestingly, we are not alone in demonstrating that p53-dependent post-transcriptional regulation of miRNA processing and function occurred only in cells expressing wild-type p53 and not in those expressing a mutant form or that the effect of the mutant differed from that of the wild-type. Suzuki et al. (2009) found that although p53 bound to and enhanced the interaction of DDX5 with the microprocessor and promoted the cleavage of pri-miRNAs to pre-miRNAs, a transcriptionally inactive mutant p53 had the opposite effect and augmented this interaction leading to a reduction in miRNA processing. In this study, the mutant forms of p53 that were investigated were described as being the transcriptionally inactive mutants C135Y, R175H and R273H (although some studies have shown gain-of-function activity for some of these mutants) rather than the S241F mutation in the DLD1 cell lines that we used.

Somatic mutations in p53 occur in around 50% of all human cancers. The majority are found between codons 125-300 corresponding predominantly with the DNA binding region, and over 80% of these involve missense mutations whereas most of those occurring outside of this zone are non-sense or frameshift mutations. Most (90%) involve single base substitutions, but base insertions and deletions are also demonstrated. Mutations may be functional, non-functional/inactivating, partially functional, or ‘supertrans’ and can cause changes in the transcriptional or transactivational properties of p53 as well as its ability to interact with other molecules such as RNA or proteins. Some mutations (e.g. N210Y) cause a loss of transcriptional activity, whereas others (e.g. R175H, D281G) activate gain-of-function (GOF) capabilities leading to the transcription of a set of genes that are not necessarily induced or suppressed by wild-type p53 but that often have net oncogenic effects. Based on work by others, the function of the S241F p53 mutant expressed in DLD1 cells used in our experiments appears to be stimulus specific. Sur et al. (2008) demonstrated that in wild-type DLD1 cells that expressed one S241F mutant allele and a one silenced allele (S214F/SIL), the DNA damaging agent 5-fluorouracil (5-FU) induced an increase in p53 but not CDKN1A. However, when the mutant S241F allele of was replaced with a wild-type allele, the level of p53 protein was reduced and p21 was induced following 5-FU treatment suggesting that the S241F mutant was transcriptionally inactive (at least in the context of CDKN1A). However, treatment with ionizing radiation did induce the differential expression of a number of genes in the S214F/SIL DLD1 cells, mainly those involved in G2/M transition or spindle assembly checkpoints, suggesting that this mutant form of p53 did regulate transcription to some degree. Our results supported this view as
we showed an increase in miR-34a expression in DLD1 cells following DOX treatment in the same manner as that demonstrated in HCT116 and RKO cells that expressed wild-type p53. Interestingly, there is some evidence in the literature that the R241F p53 mutant is not capable of forming interactions with other proteins which could affect its ability to regulate cellular processes in a non-transcriptional manner. The tumour suppressor, BARD1, and the mRNA 3’ processing factor, cleavage-stimulation factor 1 (CstF1) interact with the C-terminus of p53 following DNA damage induced by ultraviolet-radiation (UV) and this inhibits the mRNA 3’ cleavage step during polyadénylation.\textsuperscript{475} Nazeer et al. (2011), elegantly demonstrated that in p53 wild-type cell lines, p53 coexisted in complexes with CstF1 and BARD1 following UV-treatment, but that in DLD1 cell lines the interaction of p53 with these proteins was significantly reduced, as was UV-induced inhibition of 3’ processing.\textsuperscript{475} Although the DNA damaging agent here was different to that used in our experiments, this paper provides support for our hypothesis that the AGO2 loading effect might be lost in DLD1 cells due to the inability of the R241F mutant to interact with the relevant proteins involved in this process, but further work is required to confirm this.

4.1.2 How does p53 mediate the remodelling of AGO2 binding to the 3UTR of different targets at specific RNA motifs?

Having demonstrated a p53-depended effect on the loading of specific miRNAs onto AGO2 following DOX treatment, we investigated whether DNA damage also modulated the binding of AGO2 on the 3’UTR of mRNA targets independently of the miRNAs regulated following this damage. We found that, in addition to inducing or suppressing AGO2 binding to the 3’UTR of target mRNAs, DOX treatment also induced a p53-dependent remodelling of this interaction that did not occur exclusively within the miRNA-binding region of the 3’UTR. Therefore, this differential interaction appeared to occur independently of mRNA targeting by AGO2-bound miRNAs and suggested that RBPs modulated by DNA damage might interact with the 3’UTR of specific transcripts and antagonize or aid their interaction with AGO2 and hence provide an alternative mechanism of post-transcriptional regulation of gene expression. Remarkably our de novo motif discovery analysis showed that this remodelling of AGO2 binding to the 3UTR of different targets was related to a specific 10nt RNA motif on the regions of the target 3’UTR that corresponded to the sites that lost or gained AGO2 binding. This suggested that DOX treatment lead to an increase or decrease in
the interaction between an unidentified RBP and mRNAs at this specific RNA motif and that this lead to a reduction or augmentation in the binding of the mRNA onto AGO2, respectively. A particular RNA motif that was highly enriched for GUC at positions 8 to 10 but deplete for G at positions 1 to 7 was specific to p53 wild-type cells, which suggested that this motif was selected for by a p53-dependent mechanism and that this process probably involved a p53-regulated RBP. Unfortunately, our TOMTOM (http://meme.nbcr.net/meme/cgi-bin/tomtom.cgi) analysis was unable to identify an RBP interacting motif that corresponded to this RNA motif and we have therefore been unable to determine with certainty which RBP may be responsible for this process. However, following an extensive literature search we were able to identify a number of potential candidates of which the CELF family binding proteins appeared the most interesting.

The CELF family comprises of 6 (CELF1-6) evolutionarily conserved RNA-binding proteins that contain 3 highly conserved RNA-Recognition Motifs (RRM), 2 of which reside in the N-terminal and 1 of which is found in the C-terminal. The RRMs confer RNA binding activity, and it has been suggested that a divergent linker domain that bridges the RRMs is crucial for functional regulation.\textsuperscript{476-480} CELF1 and 2 are ubiquitously expressed in humans and mice, whereas CELF3-6 are found almost exclusively in the nervous system.\textsuperscript{476-480} CELF1 and 2 have been shown to play essential roles in post-transcriptional gene regulation through their ability to bind to mRNAs and affect splicing, deadenylation, mRNA translation and mRNA decay,\textsuperscript{476,481-483} but we could not find any studies that demonstrated a role for these proteins in the regulation of mRNA loading onto the RISC complex. Initial studies illustrated that CELF1 and 2 bound to mRNAs containing (CUG)\textsubscript{8} repeat sequences in their 3'UTRs,\textsuperscript{484,485} and subsequent work revealed that whilst both bound preferentially to GU-rich RNA sequences,\textsuperscript{486} CELF1 also bound to UG repeats in yeast.\textsuperscript{487} RIP-ChIP approaches in HeLa cells, human T cells, and mouse myoblasts demonstrated an over-representation of the GRE sequence, UGUUUGUUUGU, and a GU-repeat sequence, UGUGUGUGUGU in the 3'UTRs of CELF1 target mRNAs and both sequences were shown to function as mRNA decay elements.\textsuperscript{488-490} Furthermore, CELF1 and 2 were demonstrated to bind to (GUC)n triplets located on a number of mRNA targets that are known to be involved in apoptosis and cell cycle pathways (e.g. CDKN1A, p27, oskar).\textsuperscript{490,491} Interestingly, CELF1 has been shown to undergo post-translational modification, mainly through phosphorylation at various sites, which regulates its ability to bind its target mRNAs. The fact that we identified an RNA motif that was highly enriched for GUC at positions 8 to 10 but deplete for G at positions 1 to
7 within mRNAs bound to AGO2 in a p53-dependent manner, and that p53 activation is known to mediate post translation modifications to many proteins suggests that CELF1 may be the RBP involved in the remodelling of mRNA binding to AGO2 that we identified. However, further experiments are again required to validate this hypothesis. A RIP-seq and PAR-CLIP approach, immunoprecipitating for CELF1 in the presence and absence of DOX in p53 wild-type and null cells, may hold the key to achieving this.

4.1.3 Feedback and feed-forward loops provide the means for p53 to survey and coordinate the cellular response to DNA damage and complex miRNA:mRNA regulatory networks enable fine-tuning of this response

Our experimental approach outlined in Chapter 3.1 allowed us to determine the miRNA-mRNA interaction networks involved in the response to DNA damage both in the presence or absence of p53, and to do so on a larger scale than previous studies have described. This allowed us to understand in greater detail how miRNAs can function as cellular rheostats through their ability to target multiple genes involved in many response pathways induced by DNA damage and can assist a cell in switching from one fate to another or to be primed to alter its response once a stimulus had resolved. Furthermore, we also showed that the miRNAs whose cellular abundance or differential loading onto AGO2 was regulated by p53, were involved in an intricate network of regulatory feedback and feedforward circuits that fine tune gene expression levels in response to DNA damage to permit the repair of DNA damage or initiation of programmed cell death.

A striking finding was that the gene networks controlled by a specific up-regulated or down-regulated miRNA or a miRNA that was loaded more onto AGO2, could simultaneously involve genes that were suppressors or promoters of the cellular process co-ordinated by that gene network. This was exemplified by the fact that some of the genes loaded more onto AGO2 following DNA damage were pro-apoptotic and others were anti-apoptotic. In the same way, some promoted DNA repair and others inhibited it. We found that whilst miR-143-3p targeted a number of anti-apoptotic genes and genes that promoted cell cycling/division such as CCND1, it also targeted pro-apoptotic genes such as BCL2 Binding Component 3 (BBC3; also known as PUMA: p53 up-regulated modulator of apoptosis) which encodes a member of the BCL-2 family of proteins. BBC3 cooperates with direct activator proteins to induce mitochondrial outer membrane permeabilization and apoptosis. It can
bind to anti-apoptotic Bcl-2 family members to induce mitochondrial dysfunction and caspase activation and is thus an essential mediator of p53/TP53-dependent and p53/TP53-independent apoptosis.\textsuperscript{493,494} This was also the case for miRNAs that were up-regulated in a p53-independent manner. For example miR-3065-5p targeted pro-apoptotic and anti-proliferative mRNAs such as BBC3 and PTEN whilst simultaneously targeting anti-apoptic and pro-proliferative mRNAs such NACC1 and PIM1.

Furthermore, a miRNA could also target multiple mRNAs involved in conflicting cellular responses to a specific stimulus. Following DNA damage, p53 usually either co-ordinates a process of DNA repair or programmed cell death. However we found that following DNA damage a miRNA could target genes that negatively regulated apoptosis and hence induced cell death, whilst also targeting genes that suppressed DNA repair or senescence and therefore promoted the repair of DNA. For example, increased loading of let-7 onto AGO2 led to increased targeting of SMC1A, a protein involved in DNA repair via its interaction with BRCA1, as well as suppression of TUSC2 and CCND1 to promote G1 arrest and apoptosis. These findings were not completely unexpected given that, following DNA damage, p53 can act to provide damaged cells the opportunity for repair whilst concurrently or subsequently promoting the death of mutated or irrevocably damaged cells.

Additionally, p53-dependent miRNAs were found to target genes that were also transcriptionally regulated in a p53-dependent manner following DNA damage. We found that cell cycle genes such as CCND1 were transcriptionally up-regulated by p53, but at the same time were co-targeted by a number of p53 up-regulated miRNAs that inhibit their translation. Furthermore, SESN1 expression was induced by p53 following DNA damage and this gene was demonstrated to be targeted by the p53-dependently up-regulated miRNA, miR-24-3p. Sestrins are known to play a role in the cellular response to DNA damage and oxidative stress, and SESN1 mediates p53-regulated inhibition of cell growth by activating AMP-activated protein kinase, which leads to an inhibition of the mammalian target of rapamycin protein.\textsuperscript{495} The story of CDKN1A in this context was even more interesting. We demonstrated (as previously shown) that p53 transcriptionally up-regulated CDKN1A expression but we also found that the mRNA corresponding to this gene was targeted by a number of p53-dependently up-regulated miRNAs such as miR-200a-3p. Furthermore, CDKN1A was also found to be targeted by miRNAs that were down-regulated in a p53-dependent manner such as miR-301a-3p and miR-301b which, in isolation, might result in an
increase in CDKN1A activity. Adding further intricacy to the regulation of CDKN1A expression, we found that miR-185-5p, whose loading onto AGO2 increased following DNA damage, targeted RBM38 (Figure 45). This RNA-binding protein is transcriptionally induced by p53 and specifically binds to the 3'UTR of CDKN1A and maintains its stability, thereby promoting cell cycle arrest in G1 following DNA damage. Interestingly, RBM38 is also required to decrease miRNA accessibility on a number of p53-induced transcripts, allowing an optimal target gene induction and cell cycle control. However, targeting of RBM38 by miR-185-5p would lead to a reduction in CDKN1A stability which could lead to a reduction in its abundance and hence a potential increase in cell cycling. Another example of a complex p53-dependent feedback loop that we identified following DNA damage involved CCND1, as mentioned above. We showed that this gene was transcriptionally up-regulated in a p53-dependent manner following DOX treatment but that it was also targeted by miRNAs that were up-regulated in a p53-dependent manner. Furthermore, we also showed that CCND1 was targeted by miRNAs that were down-regulated in a p53-dependent manner. This suggested that the activity and function of CCND1 following DOX induced DNA damage was regulated by a complex network of feedback and feedforward loops that acted to fine tune expression of this gene and ensure it functioned appropriately at different stages of the DNA damage response.

![Complex network involved in p53-mediated regulation of CDKN1A](image_url)

Figure 48: Complex network involved in p53-mediated regulation of CDKN1A
We found further complex regulatory loops involving genes targeted by miRNAs that were regulated in a p53-independent manner following DOX treatment. We demonstrated that PTEN was a target of miR-3065-5p and it has been shown that this protein functions as a tumour suppressor by dephosphorylating and hence antagonizing components of the PI3K-AKT/PKB signalling pathway, thereby suppressing cell proliferation. However, PI3K-AKT/PKB signalling also promotes the phosphorylation and movement of Mdm2 into the nucleus where it down-regulates p53.\textsuperscript{30,31} So, by inhibiting the activation of Akt, PTEN restricts Mdm2 to the cytoplasm and promotes p53 function. Furthermore, following DNA damage p53 acutely induces Mdm2 transcription, providing damaged cells the opportunity for repair, but subsequently induces PTEN transcription by interacting with a p53-binding element directly upstream of the PTEN locus thereby favouring the death of mutated or irrevocably damaged cells.\textsuperscript{421,496} In aggregate these results suggested the existence of a complex interaction network involving p53, Mdm2, PTEN and miR-3065-5p.

Another interesting finding from the joint analysis of the PARCLIP and RNA-seq data was that in some cases the interaction network for certain miRNAs contained a combination of both up-regulated and down-regulated genes as classified by the RNA-seq, whilst in other cases the miRNA in question only targeted genes shown to be down-regulated. This suggested that the mRNA targeting and translational repression mediated by certain miRNAs was stronger than others and that in some cases the miRNA could be the most dominant factor in suppressing the expression of certain genes, even over that of transcriptional regulation. For example, let-7 targeted genes such as CCND1 whose abundance in the RNA-seq analysis was shown to be increased following DOX treatment, but that it also targeted genes such as Polymerase DNA directed lambda (POLL), HMGA2 and pyruvate dehydrogenase phosphatase catalytic subunit 2 (PDP2) whose expression decreased. Conversely, miR-3065-5p whose expression was so strongly induced following DNA damage in the RIP-seq was shown only to target genes that were down-regulated in the RNA-seq. This suggested that the 3-fold increase in the level of miR-3065-5p that occurred following DOX treatment was sufficient to ensure that all the expression of all the genes it targeted were suppressed.

These findings added further complexity to our understanding of the mRNA;miRNA interactome following DNA damage and demonstrated that the effects of miRNA-mediated post-transcriptional regulation of gene expression can depend on a balance between the
magnitude of the variation in the expression of miRNAs that co-target that specific gene through direct or indirect mechanisms. Taken together, these findings suggested that following DNA damage miRNAs are involved in multiple feedback and feed-forward loops and that they have the ability to fine-tune gene expression and regulate fluctuations in, and maintain the fluidity of, the balance between suppression and promotion of cellular response mechanisms following such a stress. Thus, following DOX induced DNA damage, oncogenes, tumor suppressor genes and miRNAs are networked through p53-dependent and p53-independent mechanisms to promote normal cell function and eliminate mutated or damaged cells.

4.1.4 DNA damage induces the expression of the GAS5-derived snoRNAs in a p53-dependent manner in colorectal cancer cell lines, and a positive correlation exists between p53 and the same snoRNAs in normal, pre-malignant and malignant colorectal tissue

Until Chang et al. (2002) first described the potential role of snoRNAs in tumorigenesis, there had been little justification for the systematic evaluation of the role of snoRNAs in this or any other pathological condition. However data are accruing that link a dysregulation in the expression of various snoRNAs to the development of a number of malignancies. As the GAS5 gene hosts ten intronic snoRNAs and a lncRNA and has been implicated in oncogenesis and the regulation of cell survival and apoptosis, and given the well documented role for p53 in the same processes, we aimed to further investigate the relationship between p53 and the GAS5 snoRNAs to gain further insight into their potential role in tumorigenesis. We found that, in colorectal cancer cell lines, the GAS5-derived snoRNAs were induced in a p53-dependent manner following DOX stimulated DNA damage but that this affect was not lost when DICER was functionally knocked down. This suggested that these snoRNAs were not processed into sdRNAs with miRNA-like function as had been suggested by a previous study, and that their role in the DNA damage response did not require them to be further processed in this manner. This implied that these snoRNAs might be involved in coordinating the p53-mediated response through their role in regulating the ribosome. snoRNAs are crucial for ribosomal function and the effective regulation of translation, and p53 is a key mediator of ribosome biogenesis especially in response to so-called nucleolar stress. Furthermore, p53 has been shown to mediate the signaling link
between ribosome biogenesis and the cell cycle. It seems logical therefore that the \textit{GAS5}-derived snoRNAs might be directly induced by p53-mediated transcription following DNA damage in order to ‘streamline’ the post-transcriptional maturation and modification of rRNAs and to ensure a more efficient translation of genes required to coordinate a response to such a stress. This theory clearly requires significant further experimental evaluation not least by proving that there is an increase in the localization of these snoRNAs to the ribosome rather than an alternative cellular compartment following DNA damage. It is possible that these snoRNAs do act at a location other than the ribosome and that they may have sdRNA type function but do not require DICER processing to enable this. A DICER-independent miRNA biogenesis pathway has been identified, whereby a pre-miRNA formed following DROSHA processing is loaded directly onto AGO and cleaved by its catalytic centre to generate an intermediate 3’-end that is then trimmed. A similar DICER-independent processing pathway may exist for snoRNAs, but again further studies are required to validate this. Whether these DNA damage induced \textit{GAS5}-derived snoRNAs simply function to accommodate an increase in gene translation at the ribosome or whether they are indeed processed to sdRNAs and have miRNA-like function, their effect on gene expression could be assessed through over-expression experiments followed by RNA sequencing. This would enable us to determine their relevance in terms of facilitating the p53-dependent response to DNA damage through the post-transcriptional regulation of gene expression.

Following on from the cell line experiments, we wished to determine whether a correlation existed between p53 levels and \textit{GAS5}-derived snoRNA expression in normal, pre-malignant and malignant human tissue samples which might have suggested a role for these snoRNAs in tumorigenesis. Interestingly, we did find a strongly positive correlation between p53 and the \textit{GAS5}-derived snoRNAs in all three tissue types and in both the micro-dissected FFPE samples (where miR-34a was used as a surrogate marker for p53) and the fresh non-micro-dissected samples. This provided further evidence in support of a role for p53 in the induction of the \textit{GAS5}-derived snoRNAs and suggested a significance for this process \textit{in vivo}. Interestingly U44 and U47 levels were found to be higher in tumour samples than in normal or pre-malignant tissue which is most likely the result of higher p53 expression in these tumours. Remarkably, and some would argue paradoxically, p53 is overexpressed in up to 50% of colorectal cancers and this has been associated with a favourable prognosis in some studies. This paradox may be explained by the fact that in many cases it is mutant rather than wild-type p53 that is overexpressed and hence the effect on cell phenotype will vary
based on the function of the mutation variant. In fact, the survival for patients with colorectal cancers expressing mutated p53, has been shown to be significantly worse than for those patients with tumours that expressed the wild-type protein and this was most striking when mutations occurred outside the evolutionarily conserved regions. We did not specifically differentiate between mutant and wild-type p53 expression in our study and it would be interesting to investigate in future work whether the same correlation exists between mutant p53 and the GAS5-derived snoRNAs in human tumour samples. It is likely that this will be dependent on the specific p53 mutation as many do retain transcriptional activity although often the gene set varies from that regulated by wild-type p53. Interestingly, in our cell line work we saw an induction in GAS5-derived snoRNA expression following DOX treatment in the DLD1 cells which are known to contain the R241F p53 mutant, and this demonstrated that this mutant form of the protein was capable of transcriptionally activating the GAS5 gene. It would also have been interesting to correlate GAS5-derived snoRNA levels with outcome in these patients, however as these samples were all taken from patients with resectable Dukes Stage A-C tumours there has not been a significant gap since their diagnosis to accrue statistically meaningful results as the data are too immature.

Another important finding from our experiments was that the GAS5-derived snoRNAs were not appropriate to be used as housekeeping genes for normalising RT-qPCR experiments that used DOX to induce DNA damage. We found that the relative expression of p53-regulated miRNAs significantly differed depending on whether GAS5-derived snoRNAs or alternative housekeeping genes such as U6 snRNA or U19 snoRNA were used to normalise results. This implies that the use of the GAS5-derived snoRNAs as normalising genes in the context of DNA damage experiments would lead to an inaccurate interpretation of the results. This is in keeping with the finding of others who have shown that in experiments involving human tumour samples, snoRNA expression was as variable as miRNA expression and that normalising miRNA PCR expression data to these snoRNAs introduced bias in associations between miRNAs and outcome.
4.1.5 The potential role of p53-regulated non-coding RNAs in cancer therapeutics

p53 is commonly dysfunctional across many tumour types either due to deletion of one or both alleles or genetic mutations and therefore provides a uniquely attractive target for drug development. However, drugs that target proteins normally inhibit rather than restore their function and therefore, like many other tumour suppressor genes, p53 is not a simple ‘druggable’ target. Furthermore, as the cellular effects of the various mutant forms of p53 can vary greatly, some having GOF and others having loss-of-function capabilities, one type of therapy that targets tumours expressing a particular form of p53 may not be suitable for those expressing an alternative mutant or deletion variant. It has therefore been challenging to develop molecules that restore wild-type p53 function to cells lacking any form of the protein, or to restore normal transcriptional activation to mutant p53 proteins, although some promising agents have been developed and complex animal models have shown that activation of the p53 response can be curative even in advanced tumours. Gene therapy approaches have also been implemented to try and restore p53 function to tumour cells and wild-type p53 delivered by adenovirus vectors is now in widespread use in China. Contusugene ladenovec (Advexin; INGN-201; Introgen Therapeutics Inc) is a replication-impaired, non-integrating, serotype 5 adenoviral vector that carries the p53 gene under the control of the CMV promoter and has consistently demonstrated a tolerability and clinical efficacy as a monotherapy or in combination with chemotherapy or radiotherapy. Additive or synergistic effects have been observed in a variety of tumour types, including squamous cell carcinoma of the head and neck, colorectal, hepatocellular carcinoma, glioma, breast, prostate, and lung cancers. Other biologic approaches include the development of antisense RNAs or siRNA that activate p53 by inhibiting the function of its negative regulators including HPV E6, MDM2 and MdmX. Furthermore, oncolytic viruses designed to replicate and selectively destroy p53 defective cells are also in the pipeline and p53-based vaccines that can activate an immune response to abnormally processed p53 in tumour cells are in the clinical trial phase. A number of small molecules that directly or indirectly activate the p53 response have also reached the clinic, of which the most advanced are the p53 MDM2 interaction inhibitors such as the Nutlins and the sirtuin inhibitors. Interestingly, Nutlin-3a has been shown to reactivate the expression of CDKN1A, PUMA, p73, E2F1 and the miR-34 family of miRNAs in p53 deficient tumour cells and induce apoptosis and senescence.
However, difficulties with designing effective p53-based therapeutics have not only centred around problems with replacing inactive p53 or reprogramming mutant p53, but in the often unpredictable response to such therapies usually in combination with DNA-damaging agents. These issues are likely to result from the very complex role that p53 plays within cells and the fact that even a small alteration in its function can have a significant effect on a finely balanced system that allows cells to switch between cell cycle arrest, senescence and DNA repair, and apoptosis. For example, the retention of wild-type p53 was demonstrated to protect breast cancers from the cytotoxic effects of some forms of chemotherapy most probably by increasing resistance to DNA damaging agents through the promotion of G2/M arrest and DNA repair rather than cancer cell death. Given the complex balance involved in the p53 response, for therapeutic purposes, it may be more useful to try to target specific elements in the pathways that p53 regulates, rather than p53 itself, in order to perhaps regulate the rheostat in favour of cell death and achieve a better response to therapy. p53-regulated small non-coding RNAs may hold the key to such a strategy and the information derived from studies like ours provide a vital source from which miRNA-based therapeutics could be selected for delivery to tumours and tip the balance in such a way as to promote cancer cell death rather than repair. Such strategies might also be employed to drive cell specific effects so as to maximise cell death in tumour cells but minimise such a response in normal tissue and avoid toxicities. miRNA-based therapeutics are becoming a reality in certain diseases and by knowing the network of genes targeted by a specific miRNA or group of miRNAs, the delivery of single or multiple miRNAs through specialised delivery systems may provide the answer to controlling the complex signalling network regulated by p53 and manipulate it effectively and provide a novel and potent form of cancer therapy.

4.1.6 Future Directions

In light of the results presented in this thesis, several mechanisms involved in the interaction between p53 and small non-coding RNAs remain to be addressed.

We describe for the first time a p53-dependent process that regulates the loading of specific miRNAs onto AGO2 following DNA damage. We hypothesise that the mechanism responsible most probably involves a direct or indirect interaction between p53 and the RISC complex. This theory is supported by our findings that this loading affect is lost in a cell line
expressing a mutant form of p53. To address this hypothesis further it would be necessary to assess:

- The presence or absence of an AGO2 loading affect in cells expressing other p53 mutants. In order to achieve this we would need to transfect pCMV-Neo-Bam plasmids containing the p53 mutants R248W, R175H or R273H (and a null plasmid) into HCT116 p53null cell lines for 48 hours and subsequently treat them with DOX or vehicle. It would then be necessary to perform an AGO2 IP experiment to yield RNA from the transfected cells and use RT-qPCR to analyse the levels of the miRNAs (e.g. let-7) that demonstrated differential AGO2 loading following DOX treatment in the p53WT cell lines. It is possible that certain p53 mutants maintain the ability to regulate AGO2 loading whilst others may not, but this remains to be shown.

- Whether p53 does in fact bind to AGO2 in the cell lines used in our experiments. To achieve this it would be necessary to perform a co-immunoprecipitation experiment to investigate the molecular interaction between p53 and AGO2. This would involve ectopically expressing AGO2 and p53 in HCT116null cells and performing an AGO2 and p53 IP with the protein derived from these cells. If there is an association between AGO2 and p53 (ad vice versa) then this would suggest an interaction between these two proteins might be responsible for the p53-dependent AGO2 loading. Furthermore, it would also be interesting to ectopically co-express mutant forms of p53 with AGO2 in the same cell lines and observe whether the association between the two proteins is lost or reduced. If this were the case it would provide an explanation as to why the AGO2 loading affect was lost in the p53-mutant cell line.

- Which region or domain in p53 is responsible for its potential association with AGO2. To achieve this is would be necessary to co-transfect Flag-tagged AGO2 into HCT116null cells with Myc-tagged p53 constructs that contain deletions in various domains including the transactivation domain, proline-rich domain, DNA-binding domain and the C-terminus. A co-immunoprecipitation with anti-Flag antibody followed by Western blotting for p53 would enable us to determine whether any particular domain was responsible for the association between AGO2 and p53.
We also demonstrated another previously undescribed process whereby DNA damage modulated the binding of AGO2 on the 3’UTR of mRNA targets independently of the miRNAs regulated following this damage. This appeared to occur independently of mRNA targeting by AGO2-bound miRNAs and suggested that RBPs modulated by DNA damage might interact with the 3’UTR of specific transcripts and antagonize or aid their interaction with AGO2. Our de novo motif discovery analysis showed that this remodelling of AGO2 binding to the 3’UTR of different targets was related to a specific 10nt RNA motif on the regions of the target 3’UTR that corresponded to the sites that lost or gained AGO2 binding. A particular RNA motif that was highly enriched for GUC at positions 8 to 10 but deplete for G at positions 1 to 7 was specific to p53 wild-type cells, but our TOMTOM analysis was unable to identify an RBP interacting motif that corresponded to this RNA motif however we hypothesise that the CELF family of binding proteins may be responsible. In order to address this further we would wish to:

- Use siRNAs to knockdown the CELF family of binding proteins in HCT116 p53\textsuperscript{WT} and p53\textsuperscript{null} cells and subsequently treat with DOX or vehicle. We would then perform protein extraction and use Western blotting to measure whether CELF knockdown led to an increase in the protein levels of specific genes that contained GUC at positions 8 to 10 but were deplete for G at positions 1 to 7. This is based on the theory that following DOX treatment, the binding of CELF proteins to the mRNA of these genes increases and hence their targeting by AGO2 reduces. This leads to an increase in the expression of these genes at the protein level. Hence, CELF knockdown would be expected to cause an increase in AGO2 targeting of these genes and hence a reduction in the expression at the protein level.

- If these above experiment demonstrates a functional effect of CELF knockdown on the protein expression levels of these genes we would then perform a RIP-seq and PAR-CLIP experiment, immunoprecipitating for CELF1 and CELF2 to further prove that this RBP binds to those genes.

We also demonstrate that DNA damage induces the expression of the \textit{GAS5}-derived snoRNAs in a p53-dependent manner in colorectal cancer cell lines, and that a positive correlation exists between p53 and the same snoRNAs in normal, pre-malignant and malignant colorectal tissue. This suggests a role for these snoRNAs in the p53 response to DNA damage but our results suggest that they may not function as sdRNAs with miRNA-like
functions as they are not processed by DICER. In order to investigate their role more fully we would aim to:

- Perform nuclear, nucleolar and cytoplasmic separation experiments in HCT116 p53\textsuperscript{WT} and p53\textsuperscript{null} cell lines following treatment with DOX or vehicle and then use RT-qPCR following RNA extraction to analyse whether a differential expression of the GAS5-derived snoRNAs exists between the different cellular components. If DOX treatment induces an increase in levels of these snoRNAs in a particular compartment, this would suggest a role for them in that particular cellular region and would help to further hypothesize what their role may be in the DNA damage response.

- Analyse our AGO2 RIP-seq data to determine whether expressed/transcribed fragments corresponding to the GAS5-derived snoRNAs are loaded onto AGO2 following DNA damage in a p53-dependent manner. The presence of such fragments on AGO2 would provide evidence for a miRNA-like role for these snoRNAs and would suggest they are processed by a DICER-independent pathway.

- Produce plasmids expressing the GAS5-derived snoRNAs individually and transflect these plasmids into HCT116 cell lines in order to overexpress their transcripts. We would then perform RNA extraction, produce a cDNA library and sequence this to determine whether these snoRNAs are involved in the regulation of gene expression, particularly with respect to p53-regulated genes.

### 4.1.7 Conclusions

p53 earned its status as ‘the guardian of the genome’ following early studies identifying its role as a potent tumour suppressor able to regulate cell growth, DNA repair and apoptosis in response to DNA damage or oncogenic activation. However, over time, it has become clear that p53 also contributes to a cells response to numerous other normal and pathological conditions, and can mediate processes involved in senescence, metabolism, macromolecular synthesis, and cell differentiation and development. Hence the complexity of its function is only just beginning to be fully appreciated. Mutations or inactivation in p53 are the most frequent abnormalities observed in cancer cells. On the other hand miRNAs and snoRNAs are often dys-regulated in cancer and we felt that by refining our knowledge of the role of these small non-coding RNAs in p53-signalling networks and by gaining a better
appreciation of the genes that they target, we would achieve a greater understanding of their role in tumorigenesis. Although it is clear that p53’s ability to transcriptionally regulate the expression of many different coding genes is key to its function, it is apparent that its ability to act as a cellular rheostat is far more complex and to only study its effect on coding gene expression at the transcriptional level would have simply scratched the surface in terms of defining its true role. Therefore we specifically aimed to investigate, on a genome wide scale, the relationship between p53 and the non-coding RNAs that it regulated. In doing so, our work enabled us to characterise the role of non-coding RNAs in the post-transcriptional regulation of gene expression in response to DNA damage and to better understand how they interacted within the complex signalling pathways that mediated this response. We determined the miRNA-mRNA interaction networks involved in the response to DNA damage both in the presence or absence of p53 and found that cell cycle genes such as CCND1 were transcriptionally up-regulated by p53, as well as being co-regulated by various p53 up-regulated miRNAs that in turn inhibited their translation. This exemplifies how our work allowed us to illustrate, in great detail, the intricate network of feed-forward and feedback loops involved in the DNA damage response and added further complexity to our understanding of this regulatory process. Furthermore, our work enabled us to appreciate better the multiple levels at which genes were regulated by p53-dependent and -independent processes in response to DNA damage. Importantly, we demonstrated a previously undescribed mechanism by which p53 controlled gene expression through its ability to selectively regulate the loading of certain miRNAs onto the RISC complex and that this was relevant in the DNA-damage response and in uncontrolled cancer proliferation. In addition, we showed that those miRNAs whose differential loading onto AGO2 was regulated by p53, were also involved in intricate networks of regulatory feedback and feed-forward circuits that fine-tuned gene expression levels in response to DNA damage to permit DNA repair or the initiation of programmed cell death. Furthermore, by a combined AGO2 RIP-Seq and PAR-CLIP approach we showed a process by which p53 modulated a reduction, induction and remodelling of AGO2 binding to the 3’UTR of different gene targets at specific RNA motifs which demonstrated another novel mechanism by which p53 could control gene expression in response to DNA damage. We also showed that the GAS5-derived snoRNAs were regulated in p53-dependent manner following DNA damage and that this relationship appeared to be relevant in human tumours as well as cancer cells lines. Although we are yet to define this further, we hope that future work will illustrate the role of these snoRNAs in the DNA
damage response and in tumorigenesis, as it is becoming clear that this class of small non-coding RNAs function as more than simple regulators of the ribosome.

p53, through the range of cellular processes that it regulates, clearly plays a vital role in protecting cells from damage that may ultimately lead to their uncontrolled growth and evolution into a malignant phenotype. However strategies employed to utilize p53 clinically, and particularly therapeutically, have been surprisingly unsuccessful and this is probably due to the complexity of p53’s function and the intricate manner by which it enable cells to oscillate between one state and another. However, some recent breakthroughs have provided much hope and the majority of these have focused on manipulating the network of molecules that mediate or regulate p53 function rather than on p53 itself. However, the response to such therapeutic agents can be unpredictable and in some cases may have undesirable affects in terms of stemming cancer growth by upsetting the balance of the p53 response inappropriately. In order to avoid this and achieve greater efficacy and predictability, we need a more sophisticated understanding of the regulatory processes involved in the p53 response and a more precise insight into how each ‘player’ contributes to the overall balance of a cells response in a context and tissue specific manner. Although our work does not provide any clear answers to these questions it does provide a novel insight into the complexities surrounding the p53 response and can act as a solid platform upon which future work can be designed. Small RNA-based therapeutics and biomarker approaches are now becoming a reality in the cancer clinic, and dissecting the importance of the relationships between small RNAs and p53 identified here, might allow a better understanding of the role of p53 in different tumour types and enable a more selective and strategic approach to selecting the drugs to most effectively manipulate this complex pathway.
References


42. Lytle JR, Yario TA, Steitz JA. Target mRNAs are repressed as efficiently by microRNA-binding sites in the 5' UTR as in the 3' UTR. *Proceedings of the National Academy of Sciences of the United States of America*. Jun 5 2007;104(23):9667-9672.


Wu GS, Burns TF, McDonald ER, 3rd, et al. KILLER/DR5 is a DNA damage-inducible death receptor gene, links the p53-tumor suppressor to caspase activation and apoptotic death. *Advances in experimental medicine and biology.* 2000;465:143-151.


Pontarin G, Ferraro P, Bee L, Reichard P, Bianchi V. Mammalian ribonucleotide reductase subunit p53R2 is required for mitochondrial DNA replication and DNA repair in quiescent...


Roe JS, Youn HD. The positive regulation of p53 by the tumor suppressor VHL. *Cell Cycle.* Sep 2006;5(18):2054-2056.


