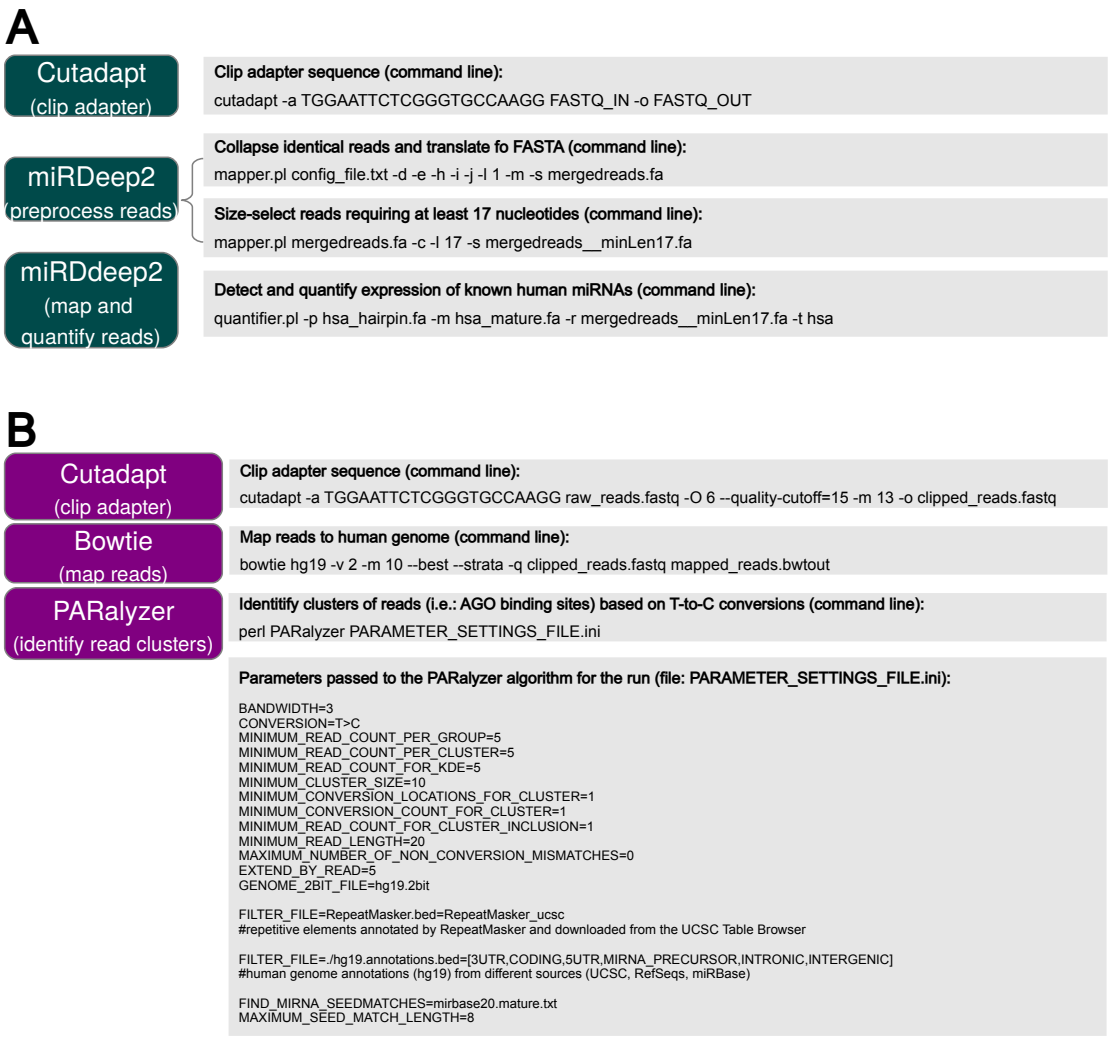
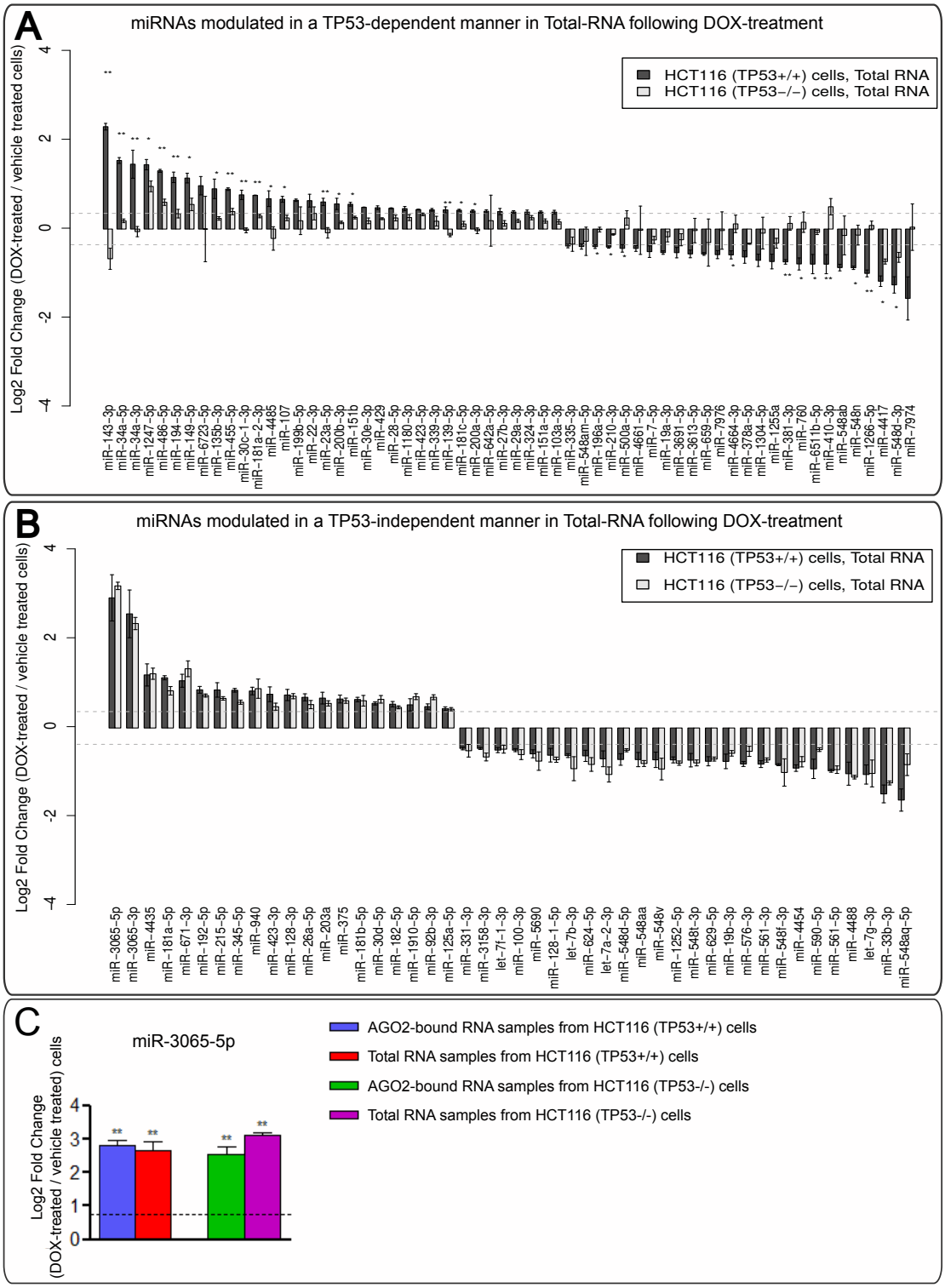
Supplemental Figures

Supplemental Figure S1



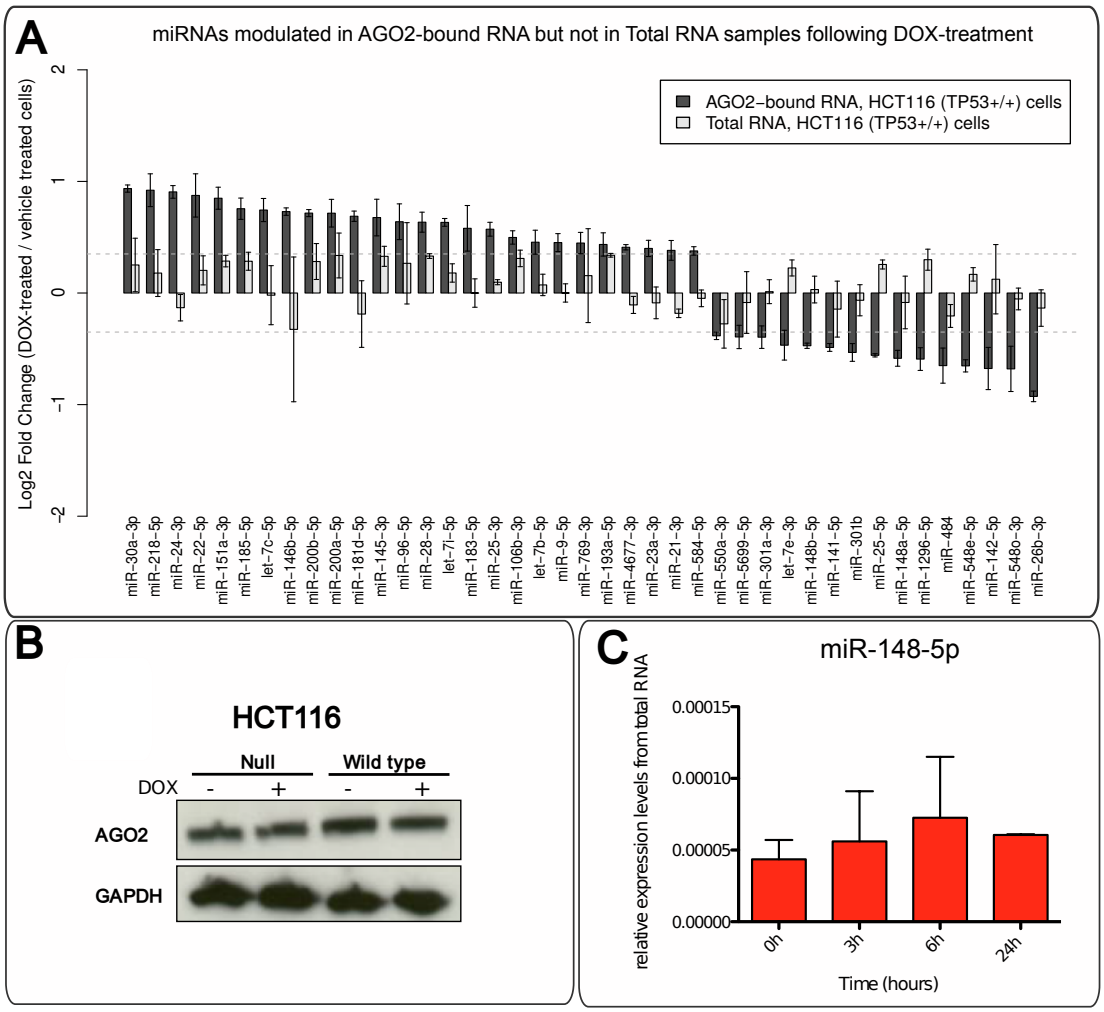
Supplemental Figure S1: Outline of data processing procedure and parameters used for the analysis of small RNA-seq, AGO2-RIP-seq and PAR-CLIP-seq data. (A), Outline of small RNA-seq and AGO2-RIP-seq data processing. Twenty-four libraries of small RNA-seq and AGO2-RIP-seq were sequenced in this study, totalling 492,467,645 sequence reads obtained by Illumina HiSeq 2000 sequencing (50nt read length, single-end). Cutadapt (https://cutadapt.readthedocs.org/en/stable/) was used to clip adapter sequences from raw small RNAs sequencing data. Modules from miRDeep2 (https://www.mdc-berlin.de/8551903/en/) were then applied to collapse clipped reads from the different samples, map them to human mature and precursor miRNA sequences and quantify the expression of miRNAs. Known human miRNA sequence annotations were taken from miRBase (http://www.mirbase.org/, release 20). A total of 363,259,138 reads were mapped to human miRNAs and used by miRDeep2 to quantify their expression in our dataset of Total RNA and AGO2-bound RNA samples, DOX- or vehicle-treated, obtained from HCT116 $TP53^{+/+}$ and $TP53^{-/-}$ cells. (B), Outline of PAR-CLIP data processing. Thirty-six libraries of AGO2 PAR-CLIP were sequenced in this study, totalling 481,293,143 sequence reads obtained by Illumina HiSeq 2000 sequencing (50nt read length, single-end). Cutadapt was used to clip the adapter sequences from PAR-CLIP sequencing data. Cutadapt was also used to remove low-quality ends of reads and discard trimmed reads shorter than 13 nucleotides. Reads were mapped to the human genome (hg19) by using Bowtie (http://bowtie-bio.sourceforge.net) (184,370,514 total mapped reads). AGO2 binding sites were identified based on T-to-C conversions and read density by using PARalyzer (https://ohlerlab.mdc-berlin.de/software/PARalyzer_85/).

Supplemental Figure S2



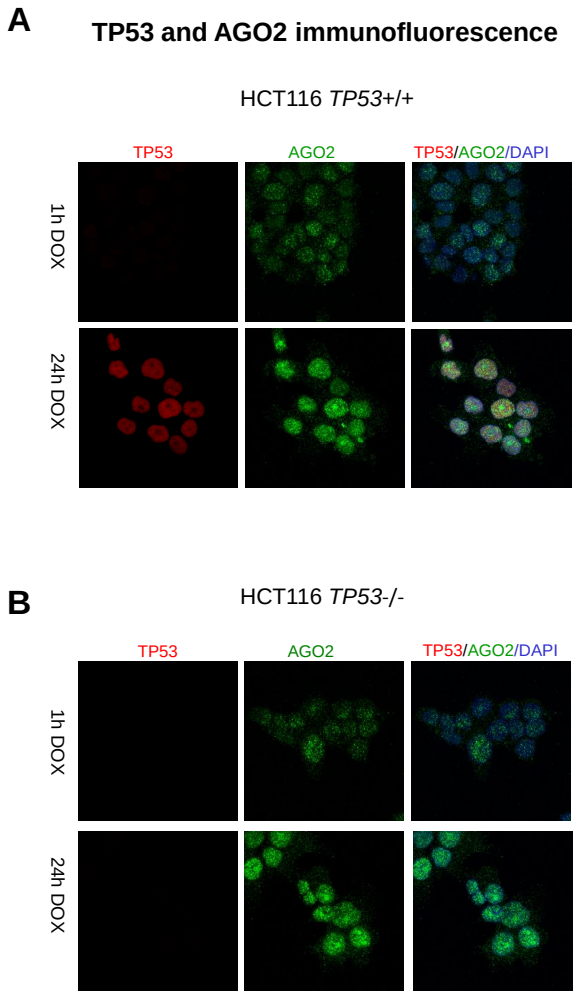
Supplemental Figure S2: A subset of miRNAs are regulated by DOX induced DSB DNA damage either dependent or independent of TP53. (A), Log2 fold change of miRNAs regulated by DOX, specifically in the presence of TP53, was measured by small RNA-Seq performed with HCT116 $TP53^{+/+}$ in comparison with small RNA-seq performed in $TP53^{-/-}$ cells. *FDR<0.05, **FDR<0.01 (Benjamini correction). Dashed lines represent the chosen cut-off threshold of log2 fold change equal to + or −0.35. (B), Log2 fold change of miRNAs regulated by DOX, independent of TP53, measured by small RNA-seq performed using HCT116 $TP53^{+/+}$ in comparison with small RNA-seq performed using $TP53^{-/-}$ cells. (C), RT-qPCR validating the regulation of miR-3065-5p by DNA damage independent of TP53. All the analysis performed here come from at least three replicates.

Supplemental Figure S3



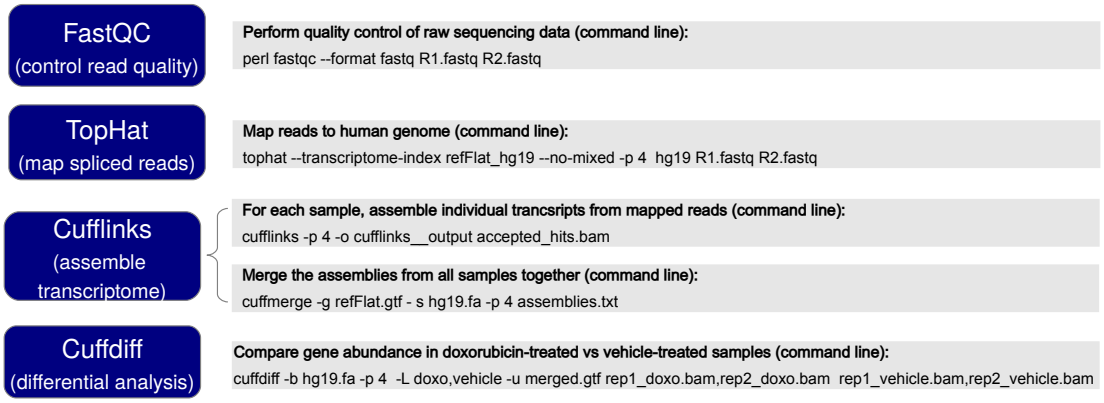
Supplemental Figure S3: TP53 regulates the binding of a subset of miRNAs onto AGO2 in response to DSB DNA damage without changes in miRNA abundance. (A), Comparison of fold change in levels of miRNA abundance observed in small RNA-seq and AGO2-RIP-seq upon DNA damage induced by DOX treatment. The miRNAs shown in the barplot exhibit TP53-dependent modulation (i.e.: positive or negative fold change $>= 0.35$ and FDR p-value < 0.05) in AGO2 association, but not in their cellular abundance, following DNA damage. (B), Western blot showing that the levels of AGO2 do not change in $TP53^{+/+}$ and $TP53^{-/-}$ HCT116 cells DOX treated or untreated. (C), Time course experiment showing no change in the abundance of miR-148-5p in HCT116 TP53+/+ cells. Three independent experiments.

Supplemental Figure S4



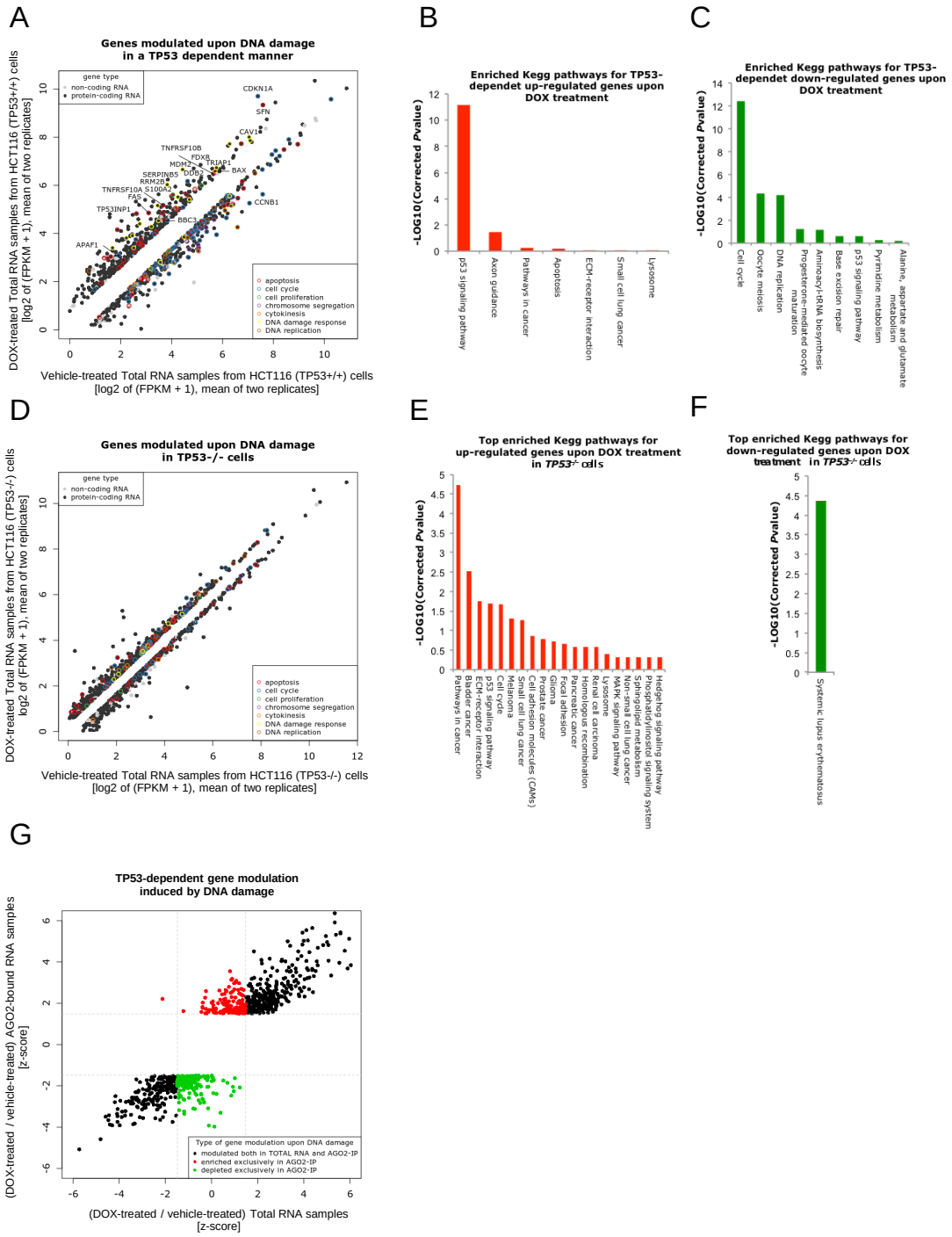
Supplemental Figure S4: AGO2 is imported into the nucleus, independently of TP53, following DNA damage. (A), Immunofluorescence assay in HCT116 $TP53^{+/+}$ cells showing that both TP53 and AGO2 are actively imported into the nucleus after 24h of DOX treatment. (B), AGO2 is also actively transported in the nucleus in HCT116 $TP53^{-/-}$ cells after 24h of DOX treatment, indicating that this effect occurs independently of its interaction with TP53. Displaying of one of three independent experiments.

Supplemental Figure S5



Supplemental Figure S5: Outline of data processing procedure and parameters used for the analysis of RNA-seq and RIP-Seq data. Sixteen libraries of RNA-seq and AGO2-RIP-seq of poly-adenylated RNAs were sequenced in this study, totaling 1,821,089,514 sequence reads obtained by Illumina HiSeq 2000 sequencing (100nt read length, paired-end). FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) was used to check the 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 (https://ccb.jhu.edu/software/tophat/) and Cufflinks (http://cole-trapnell-lab.github.io/cufflinks/) according to a published protocol. In total, 1,117,190,551 reads were successfully mapped to human genome by TopHat. Analysis of differentially expressed genes following the induction of DNA-damage through DOX treatment was assayed by using CuffDiff (http://cole-trapnell-lab.github.io/cufflinks/cuffdiff/) for the following contrasts: [i.] Total RNA samples from $TP53^{+/+}$ [ii.] Total RNA samples from $TP53^{-/-}$ cells; [iii.] AGO2-bound RNA samples from $TP53^{+/+}$ cells; [iv.] AGO2-bound RNA samples from $TP53^{-/-}$ cells.

Supplemental Figure S6



Supplemental Figure S6: Signature of DNA damage response obtained by high-throughput sequencing of poly-adenylated RNAs. (A), Scatterplot of gene expression levels in HCT116 $TP53^{+/+}$ cells for genes showing TP53-dependent modulation based on RNA-seq in DOX-treated versus vehicle-treated cells. Genes annotated to crucial functional categories in DNA damage response are highlighted with color-codes specified in the bottom-right inset. Labelled genes represent TP53 effectors from the Pathway Interaction Database (http://pid.nci.nih.gov/). Molecular gene type (coding or non-coding) is indicated in the top-left inset. (B & C), KEGG pathways enrichment analysis based on DAVID annotations (http://david.abcc.ncifcrf.gov/) for lists of up-regulated (B) and down-regulated (C) genes plotted in (A). (D), Scatter plot similar to (A), but relative to gene expression levels of modulated genes in HCT116 $TP53^{-/-}$ cells. (E & F), Similar to (B & C), but for lists of up-regulated (E) and down-regulated (F) genes plotted in (D). (G), Scatter plot of TP53-dependent modulation of AGO2-association observed in AGO2-RIP-seq (y-axis) versus TP53-dependent modulation of gene expression levels observed in RNA-seq (x-axis) induced by DOX treatment in HCT116 $TP53^{+/+}$ cells. Different patterns of gene modulation (i.e. exclusively post-transcriptional or also transcriptional) are color-coded as indicated in the bottom-right inset.

Supplemental Figure S7

A

Apoptosis related genes that increase in association with AGO2 following DNA damage

GO:0043066 (negative regulation of apoptosis)

OFFICIAL_GENE_SYMBOL	GENE NAME	Related Genes	Species
LOC646626	B-cell CLL/lymphoma 10; hypothetical LOC646626	RG	Homo sapiens
AVEN	apoptosis, caspase activation inhibitor	RG	Homo sapiens
ARNT2	aryl-hydrocarbon receptor nuclear translocator 2	RG	Homo sapiens
CIAPIN1	cytokine induced apoptosis inhibitor 1; cytokine induced apoptosis inhibitor 1 pseudogene	RG	Homo sapiens
HMOX1	heme oxygenase (decycling) 1	RG	Homo sapiens
NRG1	neuregulin 1	RG	Homo sapiens
NME5	non-metastatic cells 5, protein expressed in (nucleoside-diphosphate kinase)	RG	Homo sapiens
PIM2	pim-2 oncogene	RG	Homo sapiens

B

DNA repair related genes that decrease in association with AGO2 following DNA damage

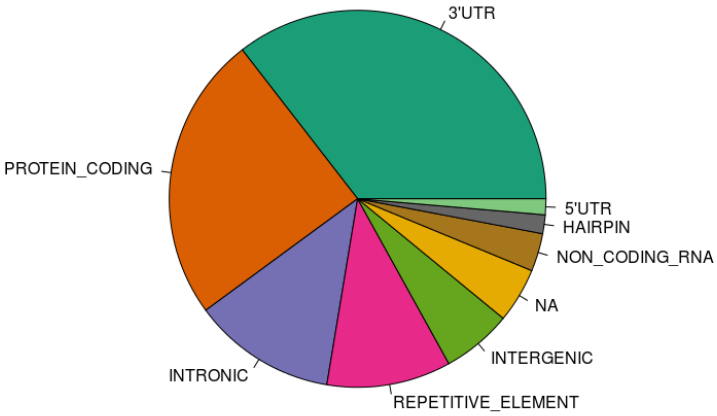
GO:0006281 (DNA repair) and GO:0006974 (response to DNA damage stimulus)

OFFICIAL_GENE_SYMBOL	GENE NAME	Related Genes	Species
OGG1	8-oxoguanine DNA glycosylase	RG	Homo sapiens
RAD21	RAD21 homolog (S. pombe)	RG	Homo sapiens
XRCC2	X-ray repair complementing defective repair in Chinese hamster cells 2	RG	Homo sapiens
XRCC4	X-ray repair complementing defective repair in Chinese hamster cells 4	RG	Homo sapiens
EME1	essential meiotic endonuclease 1 homolog 1 (S. pombe)	RG	Homo sapiens
ERCC8	excision repair cross-complementing rodent repair deficiency, complementation group 8	RG	Homo sapiens
SMC1A	structural maintenance of chromosomes 1A	RG	Homo sapiens
HAUS7	three prime repair exonuclease 2; HAUS augmin-like complex, subunit 7	RG	Homo sapiens
TOP2A	topoisomerase (DNA) II alpha 170kDa	RG	Homo sapiens

Supplemental Figure S7: A subset of genes important for the DNA damage response are either more associated or less associated with AGO2 following DOX treatment. (A), Table from Gene ontology (GO) annotation analysis performed by using DAVID (http://david.abcc.ncifcrf.gov/) identifying 8 genes that negatively regulate apoptosis (GO terms GO:0043066), which are more associated with AGO2 (more repressed) upon DNA damage. (B), As in (A) but identifying 9 genes that are involved in DNA repair and response to DNA damage stimulus (GO terms GO:0006281 and GO:00069874, respectively), which are less associated with AGO2 (their repression is released) upon DNA damage.

Supplemental Figure S8

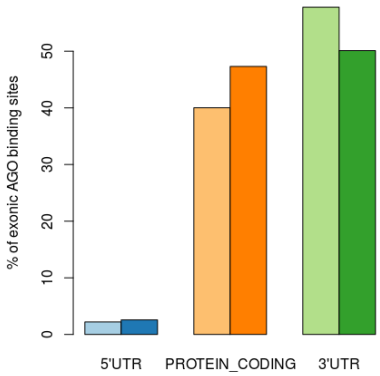
A



B

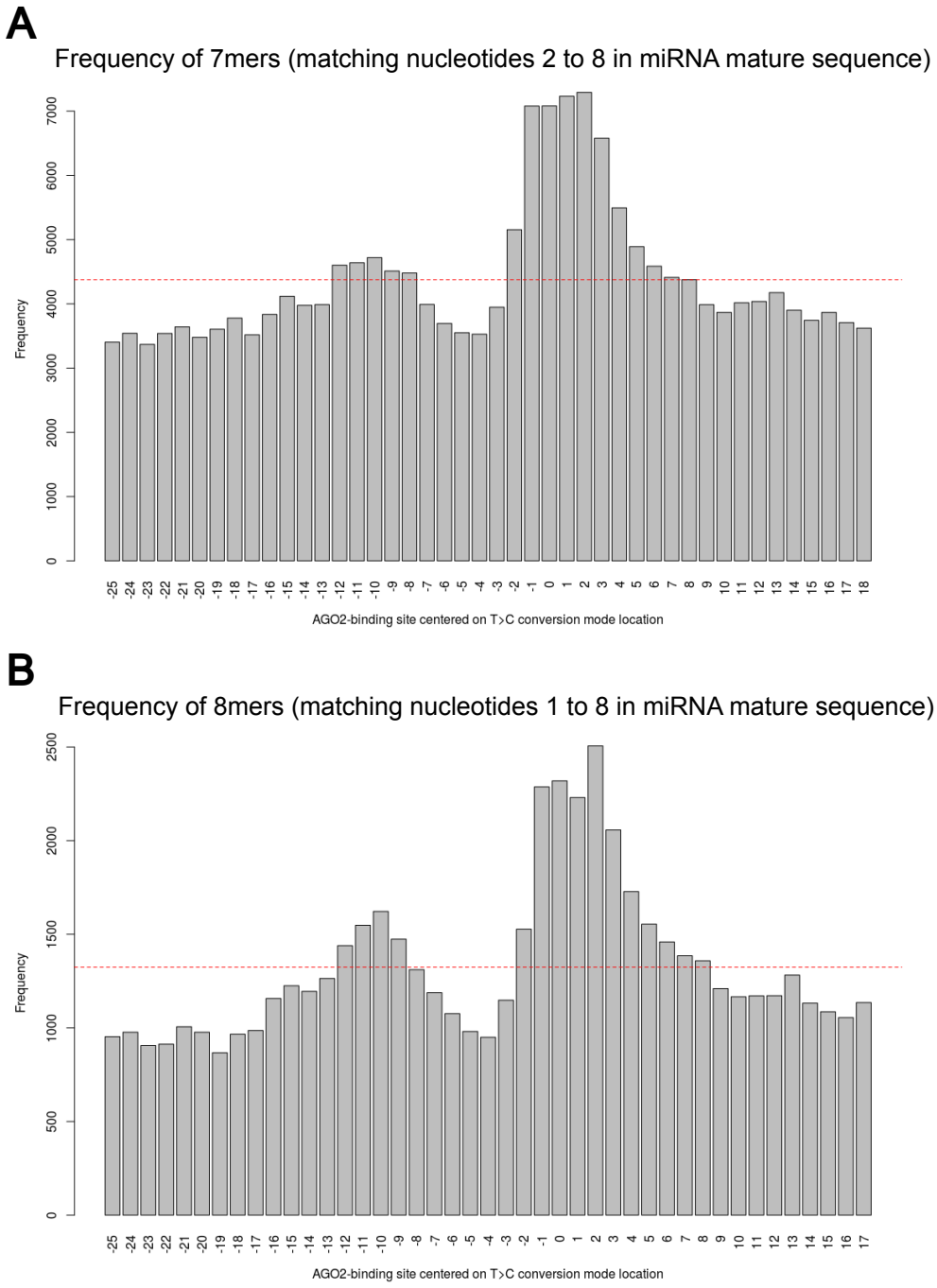
	A	T	C	G
CDS	26.18%	21.78%	25.51%	26.52%
3′UTR	27.52%	29.68%	21.32%	21.43%
5′UTR	22.75%	22.08%	27.15%	28.03%

C



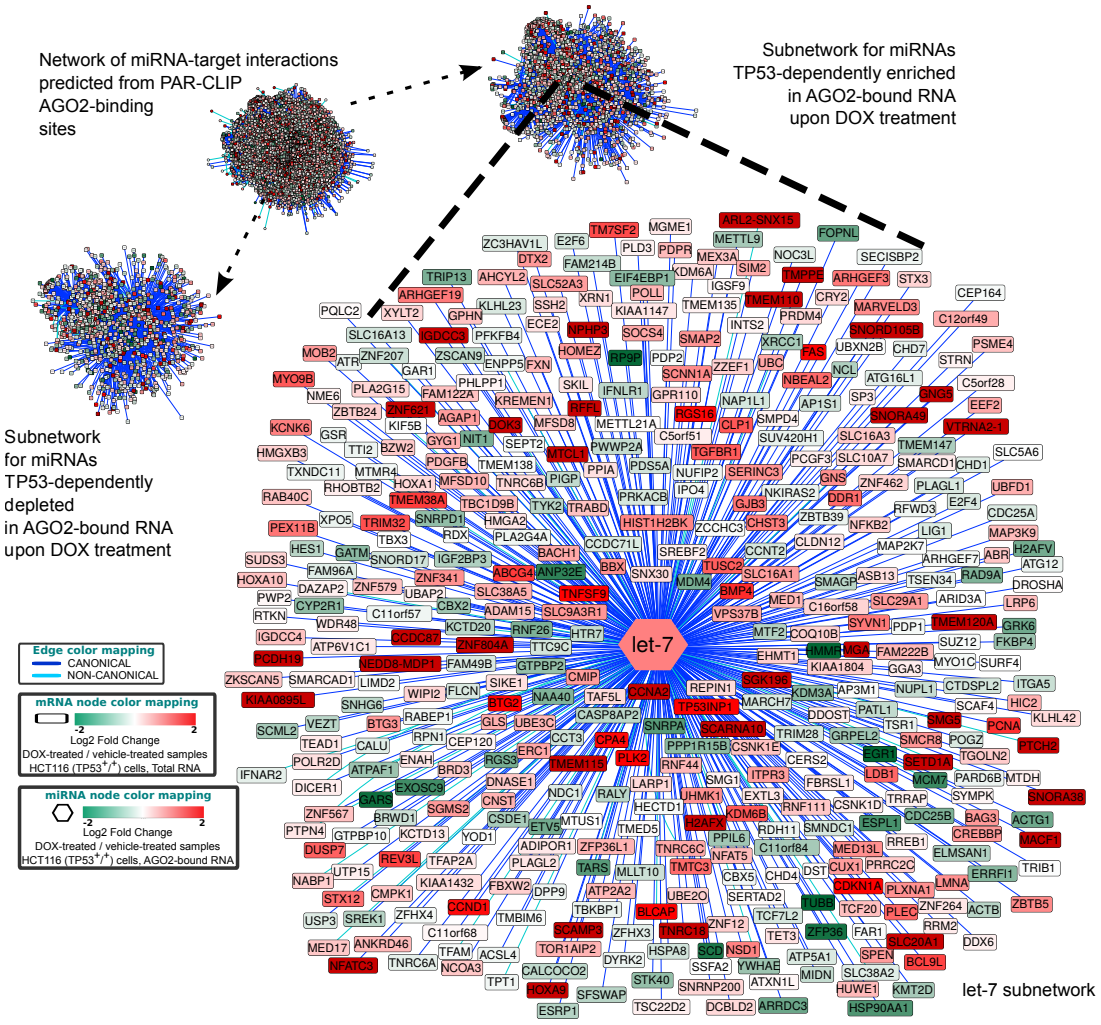
Supplemental Figure S8: Genomic characteristics of AGO2 binding sites. (A), Pie chart indicating relative abundances of genomic regions (3'UTRs, 5'UTRs, coding and non-coding exons, introns, intergenic regions, repetitive elements and miRNA hairpins) overlapped by PAR-CLIP clusters (AGO2 binding sites) identified by PARalyzer (https://ohlerlab.mdc-berlin.de/software/PARalyzer_85/) analysis. (B), Table listing nucleotide frequencies (% of As, Cs, Ts and Gs) in human CDSs, 3'UTRs and 5'UTRs based on RefSeq annotations. (C), Frequency of AGO2 binding sites (%) landing in different categories of exonic site (5'UTR, CDS, 3'UTR), before (lighter shading) and after (darker shading) adjustment for nucleotide frequencies, to correct for the propensity of PAR-CLIP clusters to land in T-rich AGO2 binding sites. After adjustment, levels of AGO2 binding to CDS and 3'UTRs are very similar.

Supplemental Figure S9



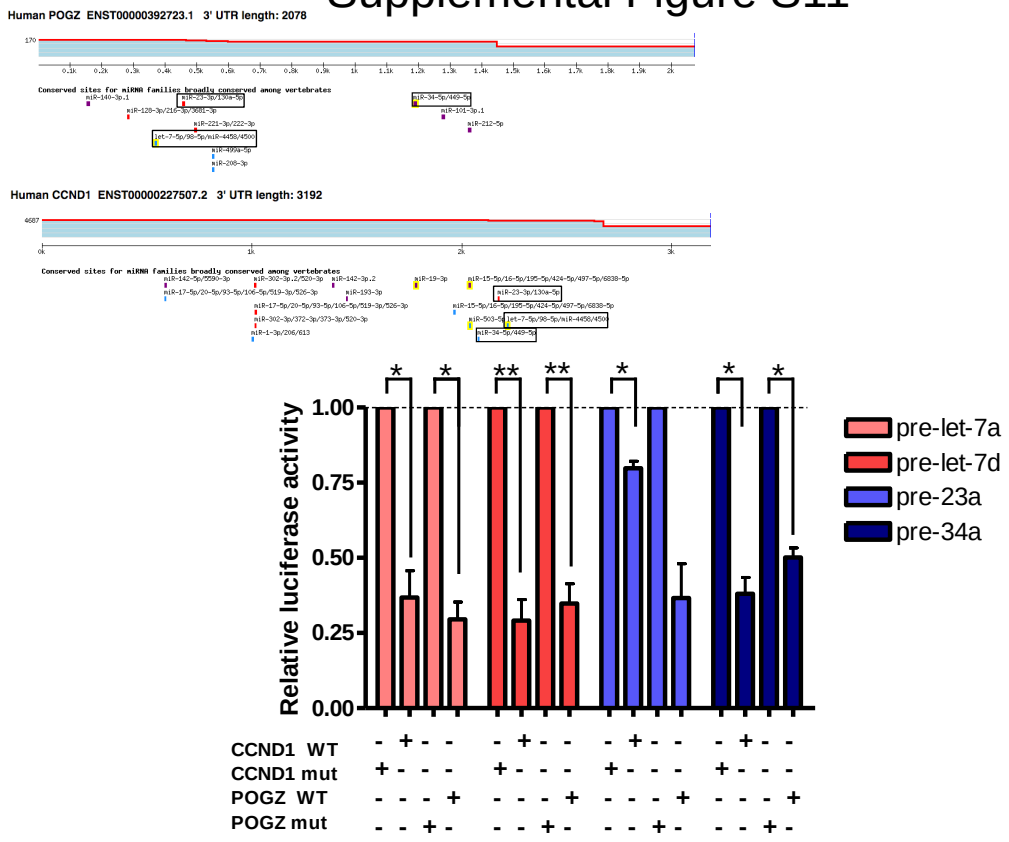
Supplemental Figure S9: Specific windows upstream to T>C conversion mode location are enriched for 7- and 8-mer miRNA seed sequences. (Top), Considering the position of maximum T>C conversion (mode location) of a cluster as 0, positions -2 to $+4$ are enriched for 7-mer (matching nucleotides 2-8 of the mature miRNA sequence) seed matches. (Bottom), Similarly, positions -2 to $+5$ are enriched for 8-mer (matching nucleotides 1-8 of the mature miRNA sequence) seed matches.

Supplemental Figure S10



Supplemental Figure S10: Global miRNA-target interaction networks driven by DNA damage. Views of the complete network of miRNA-target interactions, and selected sub-networks therein, based on PAR-CLIP AGO2-binding sites and their mapping to miRNAs. Network were visually rendered by using Cytoscape (http://www.cytoscape.org/). The complete network accounts for both canonical and non-canonical (i.e. involving or not-involving perfect complementary to the miRNA seed region, respectively) predicted interactions differently color-coded as indicated in the 'edge color mapping' inset. Different shapes identify miRNA and target gene nodes in the network, as indicated in the 'node color mapping' inset.

Supplemental Figure S11

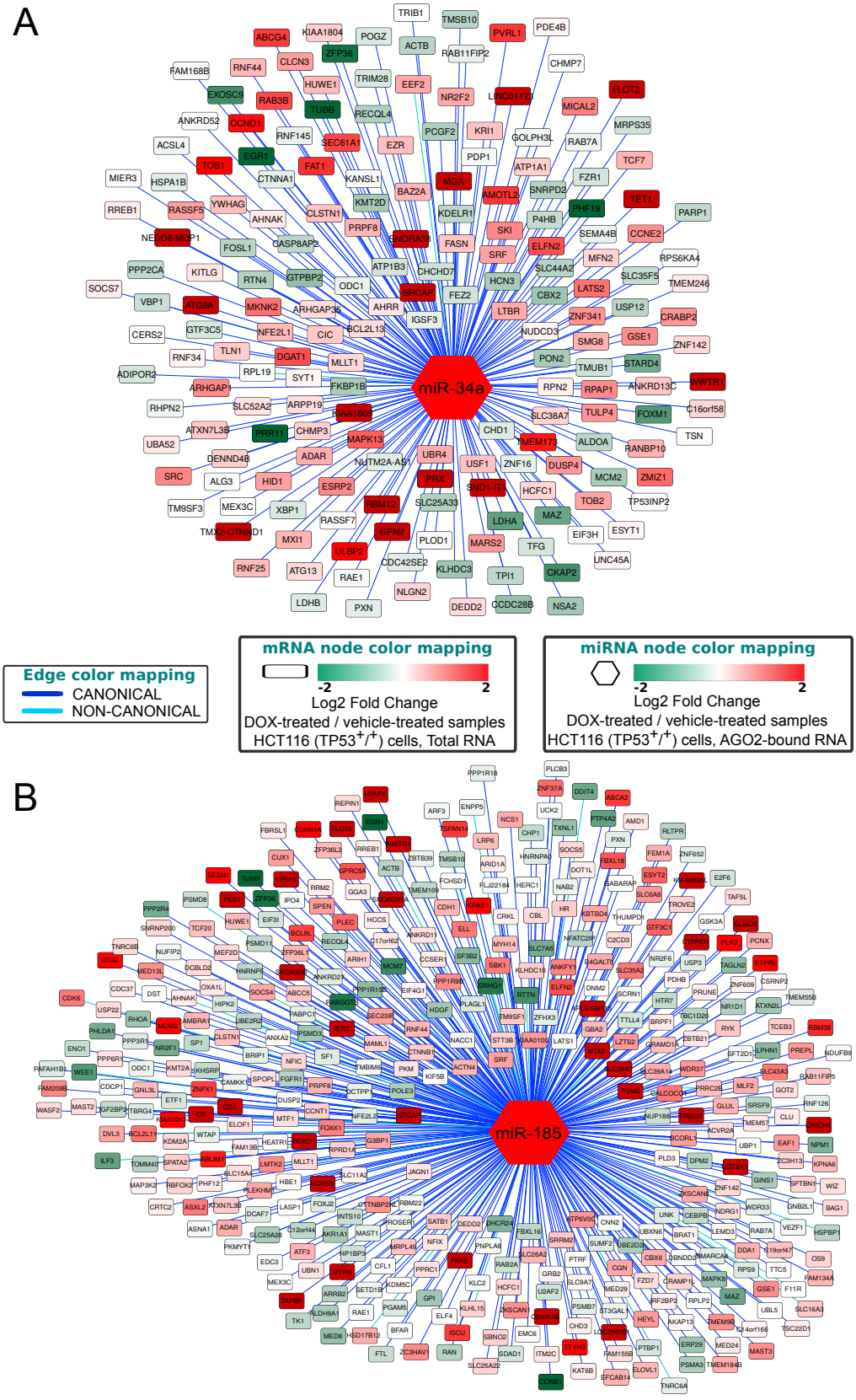



```
CCND1 3'UTR

5' ...GAUCAAUUUUGACUUAAUGUGAU...
                  |||||||
3'       CCUUUAGGGACCGUUACACUA      miR-23a
                  ||    ||
5' ...GAUCAAUUUUGACUUAAACAGAU...   23a mut


5'  ...UUUACAAUGUCAUAUACUGCCAU...
                  ||||||
3'      UGUUGGUCGAUUCUGUGACGGU      miR-34a
                  ||    |
5'      UUUACAAUGUCAUAUACACGCAU     34a mut


5'  ...GUUGCUGUUUCACAAUACCUCAU...
                  ||||||
3'      UUGAUAUGUUGGAUGAUGGAGU      let-7a/d
                  ||    |
5'  ...GUUGCUGUUUCACAAUAGGACAU... let-7a/d mut
```

```
POGZ 3'UTR

5' ...UAUCUAAUCAUUGUUAAUGUGAG...
          ||||       |||||||
3'       CCUUUAGGGACCGUUACACUA      miR-23a
          ||||       ||    ||
5' ...UAUCUAAUCAUUGUUAAACAGAG...   23a mut


5'  ...UGAAUCUUUAUUUUUCACUGCCA...
                    |||||||
3'       UGUUGGUCGAUUCUGUGACGGU     miR-34a
                    ||    ||
    ...UGAAUCUUUAUUUUUCAGACCCA... 34a mut


5' ...UUUUGUUUCCCCUUUUACCUCAA...
                    ||||||
3'    UUGAUAUGUUGGAUGAUGGAGU      let-7a/d
                    ||    |
5' ...UUUUGUUUCCCCUUUUAGGACAA... let-7a/d mut
```

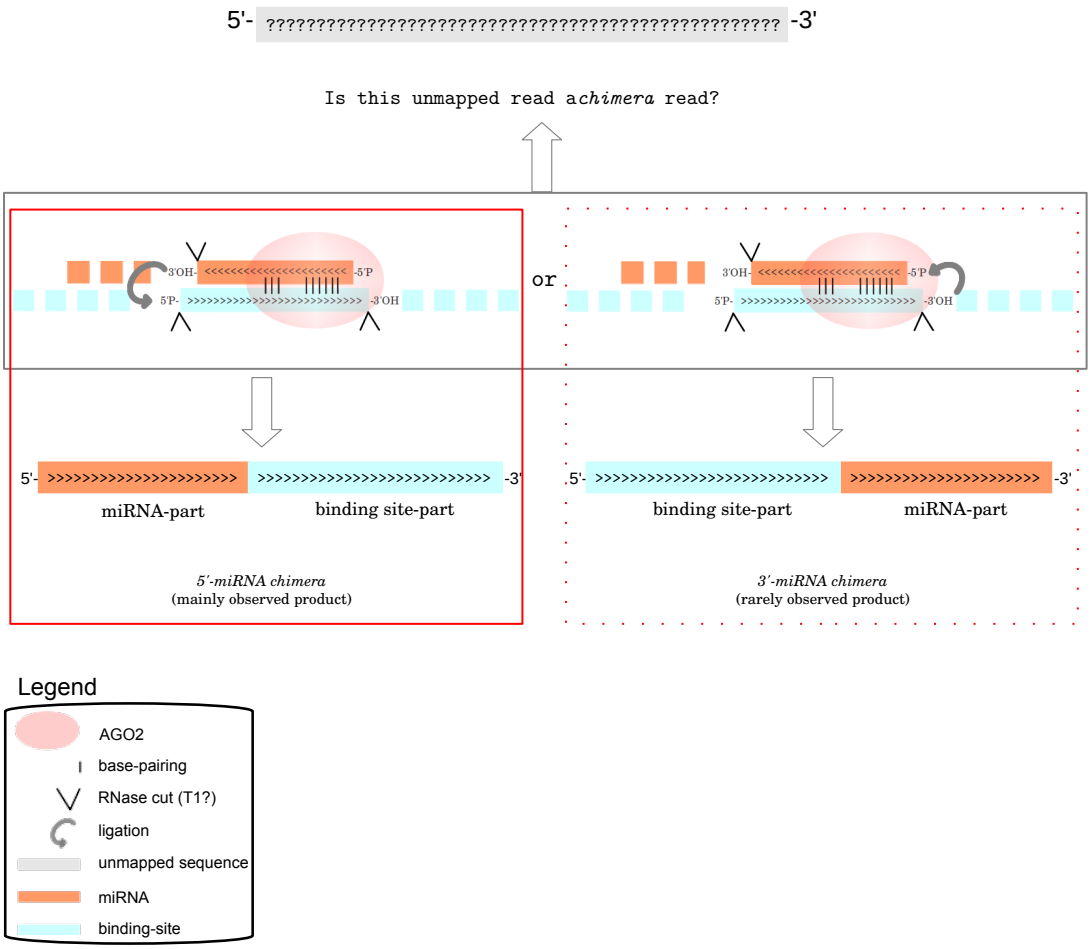
Supplemental Figure S11: *CCND1* and *POGZ* are direct targets of let-7 family members, miR-23a and miR-34a. (Top), TargetScan mapped locations of conserved miRNA binding sites on 3'UTRs of *CCND1* and *POGZ* are shown. (Middle), Relative luciferase activity levels were measured 48h after co-transfection of HCT116 *TP53*$^{+/+}$ cells with *CCND1* or *POGZ* 3'UTR-constructs (pLightSwitch_3UTR GoClone vectors from SwitchGear Genomics) and the indicated miRNA precursors. (Bottom), Locations and identities of mutations in 3'UTR of *CCND1* and *POGZ* for target sites of miR-23a, miR-34a and let-7 family members. Data represents the mean of three independent experiments ± s.e.m.

Supplemental Figure S12



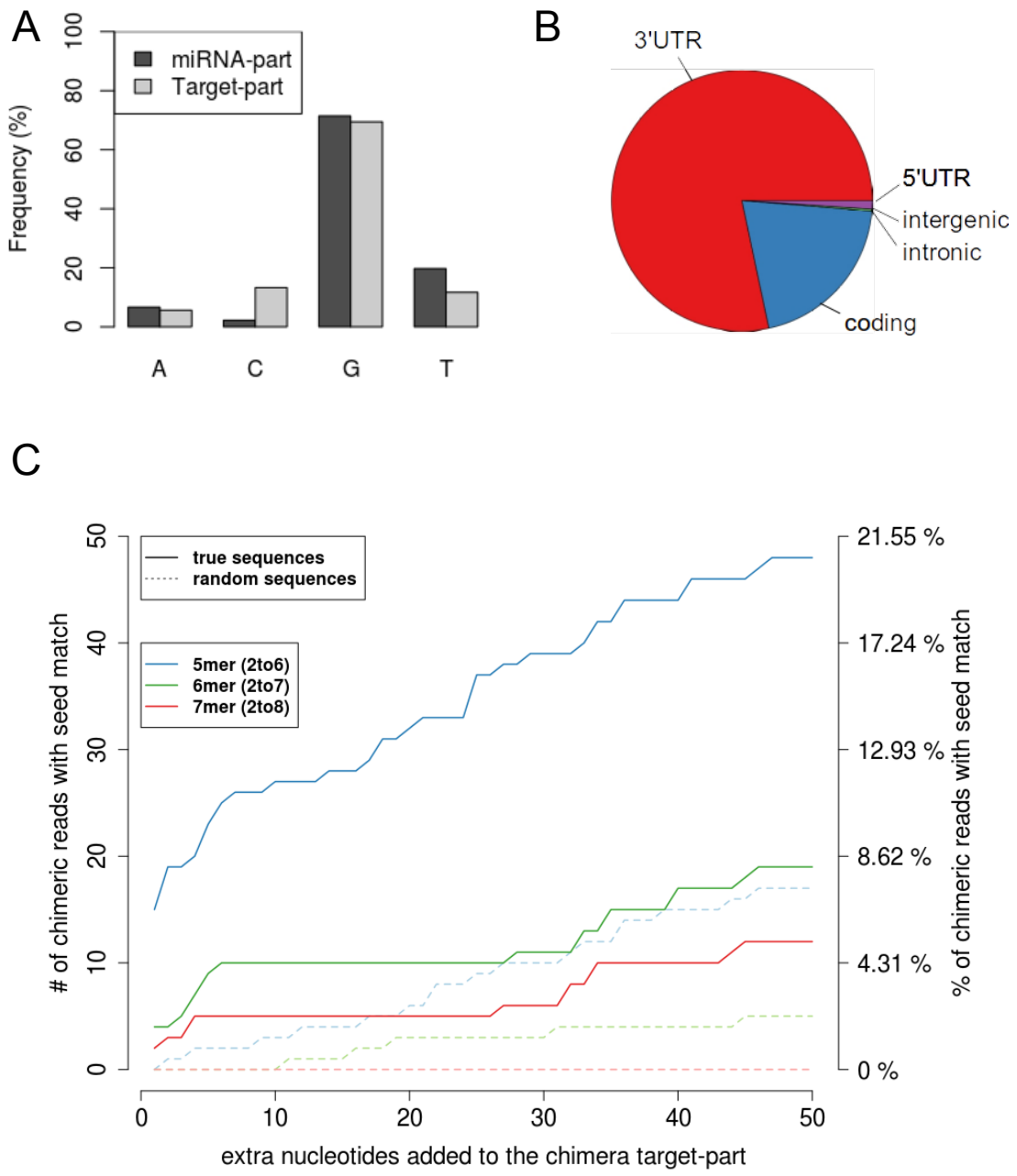
Supplemental Figure S12: Sub-networks of miR-34a and miR-185 predicted targets. (A), Sub-network of miR-34a's first neighbors nodes (i.e. miR-34a predicted targets) extracted from the complete PAR-CLIP network of miRNA-target interactions (Supplemental Figure S9 and Supplemental Table S5) inferred from mapping of the AGO2-binding sites to targeting miRNAs. The 'edge color mapping' inset specifies different colors adopted in the network for canonical and non-canonical miRNA recognition elements (MREs). Different shapes identify miRNA and target gene nodes in the network, as indicated in the 'node color mapping' inset. (B), As previously stated, but for the miR-185 subnetwork of predicted functional interactions.

Supplemental Figure S13



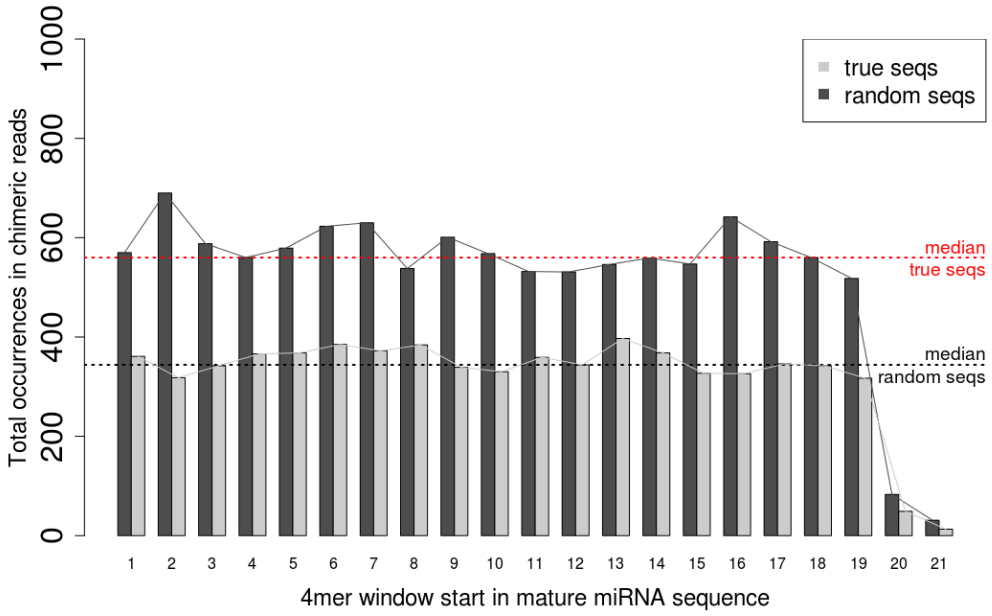
Supplemental Figure S13: Analysis of miRNA-target site interactions inferred from the computational analysis of chimic reads in identified in PAR-CLIP-seq data. Cartoon illustrating the analysis of unmapped reads in our PAR-CLIP data to unveil chimera reads (ie: fusion reads composed of part of the mature miRNA and part of its binding site sequence).

Supplemental Figure S14



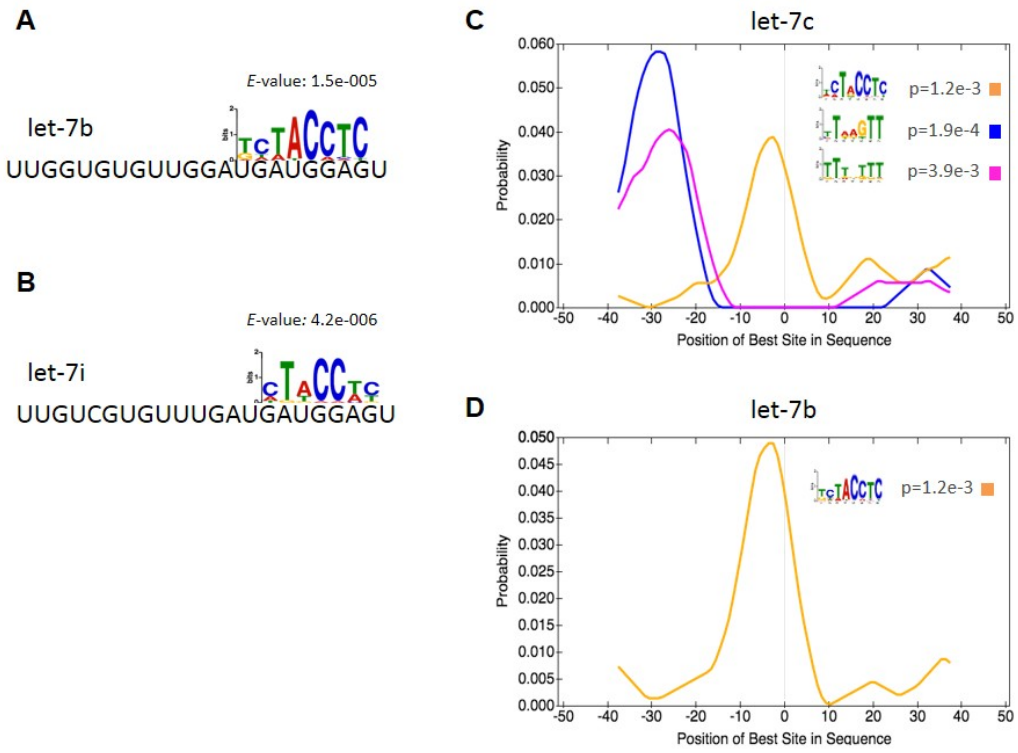
Supplemental Figure S14: Analysis of chimeric reads from PAR-CLIP libraries. (A), Analysis of nucleotide frequency at the last nucleotide preceding the miRNA- to target-part fusion point in chimeric reads confirmed previously reported enrichment for a guanosine. Abbreviations: A=Adenosine, C=Cytosine, G=Guanosine, T=Thymidine. (B) Pie chart indicating relative abundances of assayed genomic regions (3'UTRs, 5'UTRs, coding and non-coding exons, introns, intergenic regions, repetitive elements and miRNA hairpins) overlapped by genuine AGO2-binding sites identified by the re-analysis of the PAR-CLIP reads that initially failed head-to-tail mapping against the human genome due to their composite nature (i.e. part of the mature miRNA and part of its binding site sequence). (C) Analysis of miRNA seed match in target sites derived from chimeric reads. The identified chimeras were analyzed for the presence, in the target-part, of a region of complementarity to the seed region (5 to 7 nucleotides, starting at position 2 in the mature miRNA sequence) of the cognate miRNA. This was done by extending target-part sequences (ranging in size 15 to 39 nucleotides) up to 50 nucleotides upstream and downstream of their mid point by filling in their known genomic sequence (ie: creating a set of target sites of fixed - 101 nucleotides - size). The frequency of chimera sequences with a canonical seed match was then analysed as a function of the number of extra-nucleotides added 3' and compared to the frequency of matches observed when target sites were extended with random sequence (ie: a random set of sequences obtained by scrambling the order of nucleotides in each sequence from the real set, yet preserving its length).

Supplemental Figure S15



Supplemental Figure S15: Match distribution for 4-mer substrings extracted from miRNA mature sequences against the collection of target sites derived from chimeric reads. Short substrings of 4 nucleotides starting at all positions in mature miRNA sequences where analyzed for complementarity in reverse orientation to the whole dataset of chimera target sites. The same analysis was repeated on shuffled target sites sequences for estimating null profiles. The analysis was performed on miRNAs found in chimeric reads obtained from PAR-CLIP libraries.

Supplemental Figure S16



Supplemental Figure S16: Collection of target mRNAs for members of the let-7 family as identified in 3'UTR chimeric reads were analysed for motif enrichment by using CentriMo (http://meme.ebi.edu.au/centrimo.html**)** (A & B), Analysis of nucleotidic frequency from the top enriched motif identified by MEME (http://meme.ebi.edu.au/meme/tools/meme) versus the mature sequence of the related miRNA sequence (let-7b and let-7i, respectively). (C & D), Analysis of positional distribution for the best match instances of each identified motif within the collection of miRNA target sites, for let-7c and let-7b respectively. Target sites of fixed - 101 nucleotides - size were obtained by extending target-part in chimeric reads 50 nucleotides upstream and downstream from their mid point using their known genomic sequence.

Supplemental Methods

Reverse transcription quantitative real-time PCR.

Reverse transcription of mature miRNAs was performed using 10ng of total RNA and the TaqMan MicroRNA Reverse Trascription Kit (Applied Biosystems). Reverse transcription reactions 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. To amplify mature 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). Quantitative real-time PCR (qPCR) was performed with an ABI Prism 7900HT sequence detection system (Applied Biosystems). Data were analysed using qBasePlus software (biogazelle). Co-Immunoprecipitation Plasmid transfected cells were treated with doxorubicin at a concentration of 0.2ug/ml for 24 hours, then washed, scraped and lysed with a buffer containing 20mMTris HCl pH 8, 137 mM NaCl, 1% Nonidet P-40, 2 mM EDTA and inhibitors of proteases and phosphatases. 10% of total lysate was removed and

kept as the input samples and the remainder used for immunoprecipitation. 3ug of appropriate antibodies were bound to sepharose beads in the presence of heparin. Lysates were pre-cleared and then incubated with the appropriate antibody-bound beads and the immunoprecipitated proteins were then washed and eluted using 2X SDS buffer without DTT. SDS PAGE was then performed.

Plasmid transfection.

pCMV-Neo-Bam p53 R175H, pCMV-Neo-Bam p53 R248W, pCMV-Neo-Bam p53 R273H, pCMV-Neo-BAM null plasmids were purchased from Addgene (kindly supplied by Bert Vogelstein). HCT 116 $TP53^{-/-}$ cells were transfected using Lipofectamine 2000 (Invitrogen) as per the manufactures instructions.

3'UTR luciferase reporter assays.

The 3'UTRs of gene targets identified in the PAR-CLIP were cloned into pLightSwitch_3UTR Go-Clone vectors (SwitchGear Genomics, Menlo Park, CA). Cells (80% confluent) were co-transfected with pLightSwitch_3UTR luciferase reporters (50 ng/well) and the relevant pre-miR (100 nmol/L) or a non-targeting 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.

Overview of high-throughput sequencing data produced in the present study.

HCT116 human colon cancer cell lines wild-type or null for TP53 were stimulated with the DNA damaging agent doxorubicin (DOX) for 24 hours and assayed with different high-throughput techniques. Deep sequencing was used to profile the abundance of small RNAs and poly-adenylated RNAs extracted from total RNA samples (RNA-seq) or from RNA samples purified from immunoprecipitated AGO2 (AGO2-RIP-seq) obtained from either $TP53^{+/+}$ or $TP53^{-/-}$ HCT116 cells.

All small RNA sequencing experiments (small RNA-seq and AGO2-RIP-seq performed for DOX-treated and vehicle-treated HCT116 cells with $TP53^{+/+}$ or $TP53^{-/-}$) were conducted in 3 biological replicates, totalling 492,467,645 sequenced reads obtained from 24 libraries (2 RNA sample types x 4 conditions x 3 biological replicates). All poly-adenylated RNA sequencing experiments (RNA-seq and AGO2-RIP-seq performed for DOX-treated and vehicle-treated HCT116 cells with $TP53^{+/+}$ or $TP53^{-/-}$) were conducted in 2 biological replicates, totalling 1,821,089,514 sequenced reads obtained from 16 libraries (2 RNA sample types x 4 conditions x 2 biological replicates).

Photoactivatable-Ribonucleoside-Enhanced Crosslinking and Immunoprecipitation followed by deep sequencing (PAR-CLIP) method (Hafner et al., 2010) followed by high-throughput sequencing (PAR-CLIP-seq) was used to identify AGO2 binding sites and infer miRNA targets. Each AGO2 PAR-CLIP experiment of the four performed (i.e. for DOX-treated and vehicle-treated HCT116 cells with $TP53^{+/+}$ or $TP53^{-/-}$) was ran in 3 biological replicates, each sequenced in 3 technical replicates, totalling 481,293,143 sequenced reads obtained from 36 libraries (4 conditions x 3 biological replicated x 3 technical replicates).

All sequencing data produced in this study have been deposited in the European Nucleotide Archive (ENA; `http://www.ebi.ac.uk/ena`) where they are accessible through the following study identifiers: PRJEB3157 and PRJEB3233, for small and long RNAs sequencing data, and PRJEB3396, for PAR-CLIP data.

Processing of small RNAs sequencing data (RNA-seq and AGO2-RIP-seq).

Read preprocessing. Raw sequences obtained from Illumina HiSeq 2000 sequencing (50nt length) of 24 libraries of small RNA-seq and AGO2-RIP-seq (492,467,645 total sequences) were cleaned of any sequence belonging to the sequencing library adapters by using cutadapt (Martin, 2011) - version 1.2.1. *Read processing.* Clipped reads were size selected to remove sequences shorter than 17 nucleotides and identical reads collapsed by using the miRDeep2 software (Friedländer et al., 2008). *Read mapping.* Collapsed reads from all samples were then mapped to the human collection of mature and precursor miRNA sequences by using the miRDeep2 software (Friedländer et al., 2008) and miRNA sequence annotations from miRBase (Griffiths-Jones, 2004) - release 20 - and the mapped reads used to quantify the expression of miRNAs. Supplemental Fig. 1A summarizes the aforementioned steps. Overall, in the small RNAs sequencing data analysis 363,259,138 reads were mapped by miRDeep to 2578 mature miRNA sequences annotated in miRBase (Griffiths-Jones, 2004) - release 20, and used to quantify the abundance of these miRNAs across our dataset of Total RNA and AGO2-bound RNA samples, DOX- or vehicle-treated, obtained from HCT116 $TP53^{+/+}$ and $TP53^{-/-}$ cells.

In order to focus the analysis on miRNAs expressed in at least one sample condition, we first restricted the dataset to miRNAs having at least six raw read counts in at least two samples [N=1322 mature miRNAs, including redundant identifiers corresponding to multiple stem-loops yielding the same mature sequence; N=1185 unique mature miRNAs]. The *DESeq* Bioconductor package (Anders and Huber, 2010) was used to estimated library size factors and normalize read counts across samples accordingly.

Analysis of miRNA modulation following DNA-damage

To find miRNAs that showed differential expression following DOX-induced DNA-damage, mean log-fold changes (mean of 3 biological replicates) were computed for all miRNAs for the following four comparator groups of DOX-treated *vs.* vehicle-treated cells: [i.] Total RNA from DOX-treated cells *vs.* Total RNA from vehicle-treated cells, from HCT116 $TP53^{+/+}$ cells; [ii.] Total RNA from DOX-treated cells *vs.* Total

RNA from vehicle-treated cells, from HCT116 $TP53^{-/-}$ cells; [iii.] AGO2-bound RNA from DOX-treated cells *vs.* AGO2-bound RNA from vehicle-treated cells,from HCT116 $TP53^{+/+}$ cells; [iv.] AGO2-bound RNA from DOX-treated cells *vs.* AGO2-bound RNA from vehicle-treated cells, from HCT116 $TP53^{-/-}$ cells;

A moderated t-test (Smyth, 2005) was used to evaluate the significance of observed modulation. To reduce the number of tests, we restricted statistical significance analysis to the top-abundant and modulated miRNAs in each given comparator group by setting arbitrary thresholds for minimum average abundance (log2-scale) and absolute log-fold change to 5 and 0.35, respectively. For comparator groups involving AGO2-bound RNA samples, we further required miRNAs to match the above abundance threshold also in the cognate total RNA samples (Supplemental Table 1). A Benjamini-Hochberg (Benjamini and Hochberg, 1995) adjusted p-value of the moderated t-test < 0.05 was taken as significant.

Lists of regulated miRNAs were further grouped based on dependency of modulation on the TP53-status, according to the following rules:

miRNAs regulated in a TP53-dependent manner: [i] significant modulation (i.e.: adjusted p-value < 0.05) in $TP53^{+/+}$ DOX-treated samples compared to untreated controls, but not in $TP53^{-/-}$ background; [ii] mean log fold-change in $TP53^{+/+}$ DOX-treated *vs* untreated samples exceeding that of $TP53^{-/-}$ background; [iii] either absolute mean log fold-change in $TP53^{-/-}$ cells below the minimum change threshold (0.35) or to a significantly different extent of modulation in $TP53^{+/+}$ compared to $TP53^{-/-}$ cells (i.e.: t-test pvalue < 0.05 when comparing log fold-change values of DOX-treated *vs* untreated samples between $TP53^{+/+}$ and $TP53^{-/-}$ HCT116 cells).

miRNAs regulated in a TP53-independent manner: [i] significant modulation (i.e.: adjusted p-value < 0.05) in $TP53^{+/+}$ DOX-treated samples compared to untreated controls; [ii] as previously stated, but for $TP53^{-/-}$ cells; [iii] comparable extent of modulation in $TP53^{+/+}$ and $TP53^{-/-}$ cells (i.e.: no significant difference when comparing log fold-change values of DOX-treated *vs* untreated samples between $TP53^{+/+}$ and TP53 knockout HCT116 cells).

Subsets of differentially expressed miRNAs falling into each of the above-listed subcategories (i.e.: *TP53-dependent* and *TP53-independent*) were evaluated separately for Total RNA and AGO2-bound RNA samples.

Processing of poly-adenilated RNAs sequencing data (RNA-seq and AGO2-RIP-seq).

Read preprocessing. The quality of processed reads was assayed using FastQC - version 0.10.1 (`http://www.bioinformatics.babraham.ac.uk/projects/fastqc/`). Raw sequences obtained from Illumina HiSeq 2000 sequencing (100nt length, paired-end) of 16 RNA-seq and AGO2-RIP-seq libraries (1,821,089,514 total sequences) were processed by using TopHat and Cufflinks and according to a published protocol (Trapnell et al., 2012). Briefly, paired-end reads were mapped to the human genome (hg19) by using TopHat (Trapnell et al., 2009) - version 2.0.9. In total, 1,117,190,551 reads were successfully mapped to the human genome. Individual trascripts obtained from mapped reads were assembled into a reference transcriptome and quantified by using Cufflinks (Roberts et al., 2011) - version 2.1.1. Supplemental Fig. 5 summarizes the above procedure.

Analysis of gene modulation following DNA-damage.

Differences in gene expression levels between DOX-treated and control samples (two biological replicates for each experiment) were compared by using the Cuffdiff tool, that is included in the Cufflinks suite of software (Roberts et al., 2011). The differential expression analysis was repeated for the same aforementioned comparator groups of DOX-treated *vs.* vehicle-treated cells performed in the miRNA sequencing data analysis (see section *Analysis of miRNA modulation following DNA-damage*). Genes included in the Cufflinks transcriptome assembly (N=29888 Cufflinks identifiers) were filtered to remove those marked by extremely low-abundance. In particular, we required a Cufflinks flag status of "OK" and fragments per kilobase of exon per million fragments (FPKM) values above zero for genes to be considered in the next analyses (Supplemental Table 2). Next we focused the analysis of differential expression on those Cufflinks identifiers annotated to a gene symbol and made the datasets unique based on gene annotation (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 of DOX-treated *vs* untreated cells, we identified lists of modulated genes based on Z-score transformed log2 fold changes of equal to or exceeding 1.5. Z-score standardization of log2 fold changes was used to be able to set a cut-off level that would account for different variance in the different groups of samples (i.e. Total and AGO2-bound RNA samples obtained from *TP53*$^{+/+}$ and *TP53*$^{-/-}$ HCT116 cells).

Lists of regulated genes were further grouped based on the dependency of modulation on TP53-status, according to the following rules:

genes regulated in a TP53-dependent manner: [i] modulated (i.e.: z-score transformed fold-change value equal to or exceeding 1.5) in *TP53*$^{+/+}$ DOX-treated samples compared to untreated controls, but not in *TP53*$^{-/-}$ background; [ii] either unchanged (i.e.: z-score transformed fold-change value below 1) in *TP53*$^{-/-}$ cells or showing in *TP53*$^{+/+}$ cells an outstanding modulation compared to that observed in *TP53*$^{-/-}$ cells (i.e.: z-score transformed fold-change value equal to or exceeding 2 in *TP53*$^{+/+}$ cells and ratio for the z-score computed in *TP53*$^{+/+}$ cells to that computed in *TP53*$^{-/-}$ cells equal to or exceeding 1.5).

genes regulated in a TP53-independent manner: [i] modulated (i.e.: z-score transformed fold-change value equal to or exceeding 1.5) in *TP53*$^{+/+}$ DOX-treated samples compared to vehicle-treated controls; [ii] same as before, but for *TP53*$^{-/-}$ cells; [iii] comparable amount of modulation in the presence or absence of wild-type TP53 (i.e.: ratio for the z-score computed in *TP53*$^{+/+}$ cells to that computed in *TP53*$^{-/-}$ cells below 1.5)

Subsets of differentially expressed genes falling into each of the above-listed subcategories (i.e.: *TP53-dependent* and *TP53-independent*) were evaluated separately for Total RNA and AGO2-bound RNA samples, as well as for upregulated and downregulated genes (Supplemental Table 2). Selected gene sublists were tested for enriched functional categories by using either the *DAVID* bioinformatic resources

(Dennis et al., 2003) or the R/Bioconductor package *clusterProfiler* (Yu et al., 2012).

Seed enrichment analysis.

A miRNA is functional when loaded into an RNA-protein complex called the RNA-induced silencing complex (RISC), which the miRNA component directs to target mRNAs via sequence-specific patterns of base pairing with partially complementary sites in their 3'-untranslated regions (3'UTRs) (Bartel, 2009). RISC targets are largely dictated by complementarity between the seed region (nucleotides 2-7 at the 5́-end of the mature miRNA sequence) of the loaded miRNA and one or more sites in the mRNA 3'UTR (Bartel, 2009). This principle of seed primacy is accounted for by the most popular target prediction algorithms (Lewis et al., 2005; John et al., 2004; Krek et al., 2005). To evaluate whether the observed remodelling of AGO-bound mRNAs following induction of DNA damage can be attributed to an association with modulated miRNAs, we used human target predictions from the Targetscan algorithm (Lewis et al., 2005) and tested each AGO-enriched (and AGO-depleted) miRNA for an enriched proportion of their predicted targets among AGO-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 (Supplemental Table 2, sheet named "all genes"). Gene redundancy in the prediction database due to alternative transcripts was removed by retaining only the first occurrence of each gene symbol after sorting the database by decreasing 3'UTR length. Results of the seed enrichment analysis are reported in Supplemental Table 4.

Processing of AGO2 PAR-CLIP data.

Read preprocessing. Illumina HiSeq 2000 sequencing (50nt length, single-end) of 36 AGO2 PAR-CLIP libraries yielded a total of 481,293,143 sequence reads. We used cutadapt (Martin, 2011) - version 1.2.1 - to remove adapter sequences from raw PAR-CLIP-seq 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. Trimmed reads were mapped against the human genome (hg19) allowing up to 2 mismatches by using

Bowtie (Langmead et al., 2009) - version 1.0.0 - and requiring mapped locations to be reported only for reads with at most 10 different genomic hits and limiting the reported output for each mapped read to its optimal mismatch-stratum. In total, 184,370,514 were successfully mapped against the human genome. PARalyzer (Corcoran et al., 2011) - version 1.1 - was used to group mapped reads and identify "clusters" (i.e.: AGO2 binding sites) based on T-to-C conversions. Supplemental Fig. 1B reports the parameters passed to PARalyzer for the runs (.ini file settings).

PARalyzer clusters were annotated by using the BEDTools software suite (Quinlan and Hall, 2010) to known genomic features obtained from different sources - RefSeq (Pruitt et al., 2014), RepeatMasker (`http://www.repeatmasker.org/`), miRBase (Griffiths-Jones, 2004), UCSC (Kent et al., 2002), Homer (Heinz et al., 2010) - and mapped to assembly hg19 of the human genome. If a cluster mapped to a genomic location that could be linked to multiple annotations, it was assigned based on the following annotation priority: 3'UTR, coding exon, non-coding exon, miRNA precursor, 5'UTR, intron, intergenic region, repetitive elements. Overall, PARalyzer (Corcoran et al., 2011) identified a total of 111,841 AGO2-binding sites across our 36 PAR-CLIP libraries (4 conditions x 3 biological replicated x 3 technical replicates). These were then grouped based on identical genomic coordinate for the maximum of T-to-C conversion (mode location), resulting in 54,256 AGO2-binding sites falling in different categories of annotated genomic regions. Specifically, we found 20,995 AGO2-binding sites falling in 3'UTRs, 19,704 in protein coding exons, 2,885 in introns, 1,829 in repetitive regions, 1,541 in intergenic regions, 1,367 in miRNA hairpin precursors, 588 in 5'UTRs and 413 in exons of non-coding RNAs. There were also 4,934 falling into currently unannotated genomic regions (Supplemental Table 5).

Inference of miRNA-Target gene interactions from PAR-CLIP data.

AGO2-binding sites identified from the PAR-CLIP data analysis were searched for putative miRNA recognition elements (MREs) responsible for AGO2 interaction with the observed target sites. To this end, we scored conserved and non-conserved short sites matching the reverse-complement sequence of canonical miRNA seeds (Lewis et al., 2005). Specifically, we scored matches to the following seed types:

"6mer" (exact match to nucleotides 2 to 7 in the mature miRNA sequence), "7mer" (exact match to nucleotides 2 to 8), "7mer-A1" (exact match to nucleotides 2 to 7 followed by an Adenine), "8mer" (exact match to nucleotides 1 to 8), "8mer-A1" (exact match to nucleotides 2 to 8 followed by an Adenine). To evaluate MRE conservation, multi-species alignment across human, mouse, rat, dog and chicken genomes for 54,256 AGO2-binding sites were extracted from Galaxy (Blankenberg et al., 2011). To lower the number of spurious miRNA seed matches, we restricted our analysis to 181 miRNAs among the top 300 abundant miRNAs in our samples and modulated upon DOX treatment at any tested condition (Supplemental Table 1). Distribution of the position of the first MRE nucleotide in the AGO2-binding site showed a clear preference to positions from -2 to +5 with respect to mode location for T-to-C conversion (Supplemental Figure 9), in agreement with previous reports (Hafner et al., 2012). With respect to AGO2-binding sites with matches to multiple miRNAs from different miRNA families (i.e.: with different nucleotides at positions 2 to 8 in the mature sequence) the given AGO2-binding site was assigned to a single miRNA (or miRNA family) by applying the following priority rule: (i) select top conserved MRE; (ii) in case of multiple MREs with same conservation, select the longest one (considering 8mer > 8mer-1A > 7mer > 7mer-1A > 6mer); (iii) in case of multiple MREs with same conservation and length, select the closest to T-to-C conversion mode location. The full set of miRNA-target interactions inferred from the above procedure based on canonical MREs is listed in Supplemental Table 5. However, since it has been shown that miRNAs can also interact with their target mRNAs beyond canonical MREs (e.g.: with partial complementary sites including G:U wobble base pairing and/or involving 3' compensatory pairing) we furthered our analysis by including recently published interactions from CLIP-Seq data (Clark et al., 2014) obtained with the CLIPSim-MC algorithm (Xia et al., 2012) and overlapping our catalogue of AGO2-binding sites. Figure 4A shows a Cytoscape (Smoot et al., 2011) rendering of the full network of putative miRNA-target interactions - including both *seed* and *seed-less* MREs - relative to 37,840 exonic AGO2-binding sites (541 found in 5'UTRs, 17,652 in protein coding exons and 19,647 in 3'UTRs), with a zoomed-in view of the sub-network for let-7 family members. These same networks

of miRNA-target interactions are listed in two separated sheets (for the full network and the let-7 family sub-network) in Supplemental Table 5.

Analysis of miRNA-target site interactions from chimera reads in PAR-CLIP data.

As per Grosswendt and coworkers (Grosswendt et al., 2014), we reanalyzed unmapped reads in our PAR-CLIP data to unveil *chimera reads* (ie: fusion reads composed of part of the mature miRNA and part of its binding site sequence) resulted from endogenous ligation reactions (Supplemental Fig. 13). To this end, we implemented a multi-step procedure based on well-established alignment tools - Blat (Kent, 2002) and Bowtie (Langmead et al., 2009) - together with ad-hoc R (R Core Team, 2014) and Perl scripts for alignment refinement and applied it to the analysis of collapsed reads from all PAR-CLIP samples that failed to align to the human hg19 genome (unmapped reads) (see section *Processing of AGO2 PAR-CLIP data*). The procedure is outlined as followed:

step 0 - collect unmapped reads from all PAR-CLIP samples: We used the FASTX-Toolkit to create a collapsed fasta dataset of unique sequences from all PAR-CLIP samples that failed to align to the human hg19 genome, each identified by a two-fields identifier in the format: [unique sequence index]-[total sequence counts]

step 1 - identification of complete and truncated miRNA sequences (miRNA part): We used the Blat (Kent, 2002) software to locally align unmapped reads against mature miRNA sequences obtained from miRBase (Griffiths-Jones, 2004) by requiring at least a shared substring of 8 nucleotides for a match to be reported (run parameters: -tileSize=8 -minIdentity=100 -minMatch=1).

step 2 - refinement and filtering of intermediate results: We selected a subset of valid chimera candidates based on: Blat match orientation (sense); match size (>=11nt); match location with respect to the candidate chimera sequence (boundary match - either at the 5' or 3' end); miRNA matched part (5'-end of the miRNA mature sequence included in the match); remaining sequence

size outside the miRNA-matching part (>=15nt); mismatches in Blat alignment (none or T>C mutation).

step 3 - identification of binding site sequences (target part): To identify the binding site section of candidate chimera sequences, the set of remaining sequence outside the chimera miRNA part were mapped against the human genome (hg19) by using the Bowtie (Kent, 2002) aligner - version 1.0.0 - allowing up to 3 mismatches to the reference sequences but discarding sequences with multiple genomic hits.

step 4 - final refinement and filtering of results: Alignment results were selected based on: mismatches in Bowtie alignment (none or T>C mutation); binding site match location in chimera read (accept only matches accounting for the whole remaining part of the chimera read outside the miRNA part); evidence supporting the observed target site across PAR-CLIP libraries (require a cumulative value of at least 5 for the normalized read counts of the given genomic interval).

step 5 - scoring of miRNA seed match: The identified chimera reads were annotated for the presence of seed matches (5 to 7 nucleotides, starting at position 2 in the mature miRNA sequence) for the cognate miRNA in the chimera target-part. To this end, target-part sequences (ranging in size 15 to 39 nucleotides) were extended up to 50 nucleotides upstream and downstream the mid point (ie: creating a set of target sites of fixed - 101 nucleotides - size) to avoid loss of genuine miRNA matching positions possibly due to RNAse activity (Grosswendt et al., 2014).

The list of miRNA-target pairs recovered from the analysis of chimera reads is reported in Supplemental Table 6.

References

Anders S and Huber W, 2010. Differential expression analysis for sequence count data. *Genome Biol*, **11**(10):R106.

Bartel D. P, 2009. Microrna: target recognition and regulatory functions. *Cell*, **136**(2):215–233.

Benjamini Y and Hochberg Y, 1995. *Controlling the false discovery rate: a practical and powerful approach to multiple testing.*, volume 57 of *Series B*. Journal of the Royal Statistical Society.

Blankenberg D, Taylor J, Nekrutenko A, and G. T, 2011. Making whole genome multiple alignments usable for biologists. *Bioinformatics*, **27**(17):2426–2428.

Clark P. M, Loher P, Quann K, Brody J, Londin E. R, and Rigoutsos I, 2014. Argonaute clip-seq reveals mirna targetome diversity across tissue types. *Sci Rep*, **4**:5947.

Corcoran D. L, Georgiev S, Mukherjee N, Gottwein E, Skalsky R. L, Keene J. D, and Ohler U, 2011. Paralyzer: definition of rna binding sites from par-clip short-read sequence data. *Genome Biol*, **12**(8):R79.

Dennis J G, Sherman B. T, Hosack D. A, Yang J, Gao W, Lane H. C, and Lempicki R. A, 2003. David: Database for annotation, visualization, and integrated discovery. *Genome Biol*, **4**(5):P3.

Friedländer M. R, Chen W, Adamidi C, Maaskola J, Einspanier R, Knespel S, and Rajewsky N, 2008. Discovering microrna from deep sequencing data using mirdeep. *Nat Biotechnol*, **26**(4):407–415.

Griffiths-Jones S, 2004. The microrna registry. *Nucleic Acids Res*, **32**(Database issue):D109–D111.

Grosswendt S, Filipchyk A, Manzano M, Klironomos F, Schilling M, Herzog M, Gottwein E, and Rajewsky N, 2014. Unambiguous identification of mirna:target site interactions by different types of ligation reactions. *Mol Cell*, **54**(6):1042–1054.

Hafner M, Landthaler M, Burger L, Khorshid M, Hausser J, Berninger P, Rothballer A, Ascano M, Jungkamp A.-C, Munschauer M, *et al*, 2010. Transcriptome-wide identification of rna-binding protein and microrna target sites by par-clip. *Cell*, **141**(1):129–141.

Hafner M, Lianoglou S, Tuschl T, and Betel D, 2012. Genome-wide identification of mirna targets by par-clip. *Methods*, **58**(2):94–105.

Heinz S, Benner C, Spann N, Bertolino E, Lin Y. C, Laslo P, Cheng J. X, Murre C, Singh H, and Glass C. K, *et al*, 2010. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and b cell identities. *Mol Cell*, **38**(4):576–589.

John B, Enright A. J, Aravin A, Tuschl T, Sander C, and Marks D. S, 2004. Human microrna targets. *PLoS Biol*, **2**(11):e363.

Kent W. J, 2002. Blat–the blast-like alignment tool. *Genome Res*, **12**(4):656–664.

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

Krek A, Grün D, Poy M. N, Wolf R, Rosenberg L, Epstein E. J, MacMenamin P, da Piedade I, Gunsalus K. C, Stoffel M, *et al*, 2005. Combinatorial microrna target predictions. *Nat Genet*, **37**(5):495–500.

Langmead B, Trapnell C, Pop M, and Salzberg S. L, 2009. Ultrafast and memory-efficient alignment of short dna sequences to the human genome. *Genome Biol*, **10**(3):R25.

Lewis B. P, Burge C. B, and Bartel D. P, 2005. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microrna targets. *Cell*, **120**(1):15–20.

Martin M, 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, **17**(1):10–12.

Pruitt K. D, Brown G. R, Hiatt S. M, Thibaud-Nissen F, Astashyn A, Ermolaeva O, Farrell C. M, Hart J, Landrum M. J, McGarvey K. M, *et al*, 2014. Refseq: an update on mammalian reference sequences. *Nucleic Acids Res*, **42**(1):D756–D763.

Quinlan A. R and Hall I. M, 2010. Bedtools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, **26**(6):841–842.

R Core Team, 2014. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Roberts A, Pimentel H, Trapnell C, and Pachter L, 2011. Identification of novel transcripts in annotated genomes using rna-seq. *Bioinformatics*, **27**(17):2325–2329.

Smoot M. E, Ono K, Ruscheinski J, Wang P.-L, and Ideker T, 2011. Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics*, **27**(3):431–432.

Smyth G. K, 2005. Limma: linear models for microarray data. In Gentleman R, Carey V, Dudoit S, Irizarry R, and Huber W, editors, *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*, pages 397–420. Springer, New York.

Trapnell C, Pachter L, and Salzberg S. L, 2009. Tophat: discovering splice junctions with rna-seq. *Bioinformatics*, **25**(9):1105–1111.

Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley D. R, Pimentel H, Salzberg S. L, Rinn J. L, and Pachter L, *et al*, 2012. Differential gene and transcript expression analysis of rna-seq experiments with tophat and cufflinks. *Nat Protoc*, **7**(3):562–578.

Xia Z, Clark P, Huynh T, Loher P, Zhao Y, Chen H.-W, Ren P, Rigoutsos I, and Zhou R, 2012. Molecular dynamics simulations of ago silencing complexes reveal a large repertoire of admissible 'seed-less' targets. *Sci Rep*, **2**:569.

Yu G, Wang L.-G, Han Y, and He Q.-Y, 2012. clusterprofiler: an r package for comparing biological themes among gene clusters. *OMICS*, **16**(5):284–287.