Supplementary Figure 1: Evolutionary conservation of the miR-29 family and CNVK-miR-29b transfection and viability data. (a) Molecular structure of 3-cyanovinylcarbazole (CNVK) and of CNVK nucleoside crosslinked to pyrimidines. The reactive CNVK bond is highlighted in red. (b) Alignment of miR-29b-3p, miR-29a-3p and miR-29c-3p sequences in selected bilaterian animals. Bold letters represent the miR-29b seed region. Red letters display deviations from the mature hsa-miR-29b sequence. The mature miRNA sequences are highlighted. Asterisks indicate conserved nucleotides in the miRNA sequence. (c) Agarose gel electrophoresis of single-stranded (sense or antisense oligos) and double-stranded (annealed) CNVK tagged RNAs. M represents double-strand low-molecular weight DNA marker. Sense oligos contain the CNVK and biotin modifications and hence migrate slower than the antisense strand. Unmodified annealed miR-29b mimic migrates faster than CNVK containing oligos on 2% agarose owing to the lack of CNVK and biotin modifications (data not shown). (d) MTT assay based assessment of NIH3T3 cell viability measured 48 hours after transfection with CNVK-miR-29b. Bars represent mean ± SEM of three independent experiments. (e) Northern blot hybridization on 20ug input RNA and corresponding pulldown RNA extracted from NIH3T3 cells transfected with CNVK-miR-29b or CNVK-scram. tRNA-Lys was used as a loading control and to evaluate non-specific transcript pulldown. Numbers indicate band intensity with respect to loading control.
Supplementary Figure 2: The CNVK-miR-29b oligo associates with endogenous AGO2 protein and suppresses endogenous target gene expression. (a) TaqMan qRT-PCR based assessment of miR-16 expression (normalized to U6 expression) in untransfected cells and cells transfected with increasing doses of CNVK-miR-29b and CNVK-scram. (b) Representative immunoblot demonstrating association of CNVK-miR-29b and CNVK-scram mimics with Argonaute2 protein in NIH3T3 cells. 1/20th volume of the irradiated cell lysate was loaded in the Input lanes. Remaining lysate was affinity purified using streptavidin magnetic beads and all of it was used for the Pulldown lanes. GAPDH is used to evaluate non-specific protein pulldown. "*" Indicates non-specific band. M refers to molecular weight marker (kDa). The antibody used for NIH3T3 cells is anti-mouse α–Ago2 antibody (ab32381, Abcam). NIH3T3 cells were transfected with (c) 10nM and (d) 40nM concentration of CNVK-miR-29b, CNVK-scram (negative control oligo), n-miR-29b or C.elegans miR-67 (negative miRNA control) oligos and gene expression values were normalised to Hprt1. Gene expression levels in CNVK-miR-29b transfected cells are plotted as a ratio of respective gene levels in CNVK-scram transfected cells (set at 1), and gene expression values in n-miR-29b transfected cells are reported as a ratio of expression values in C.elegans miR-67 transfected cells (set at 1). Irf6 is used as non-target control gene. * P<0.05 compared to respective negative control oligos; two-tailed student’s t-test. (e) qRT-PCR analysis of two previously reported miR-29b target genes (Col1a1 and Dnmt3a) in NIH3T3 cells transfected with 10nM concentration of CNVK-miR-29b or CNVK-scram oligonucleotides. Transfected live cells (“intact”) or cell lysates (“lysed”) were UV irradiated for 10 minutes and gene expression in pulldown samples was plotted as a percentage of gene expression in total input sample. Bars represent mean ± SEM of two independent experiments. (f) qRT-PCR analysis of previously reported miR-29b target gene expression in NIH3T3 cells transfected with 10nM CNVK-miR-29b, bio-miR-29b or CNVK-scram oligonucleotides in non-UV crosslinked (0 minute) samples. * P<0.05, ** P<0.01, *** P<0.001 compared to other transfected cells; one-way ANOVA, followed by Tukeys post-hoc test. Bars represent mean ± SEM of three independent experiments.
Supplementary Figure 3: Comparison with TargetScan predictions, exploration of known target and non-target genes of interest and conservation of miR-29b targets between human and mouse. (a - b) Number of TargetScan predicted miR-29b target genes identified using the \textsuperscript{CNV}K approach compared with number observed by chance alone in mouse and human, respectively. For each calculation we generated 10000 random sets of expressed genes of the same length as detected in each category, and reported the number of miR-29b targets of the corresponding class observed in each set. Comparisons were carried out with any expressed genes with a TargetScan prediction, those with an “8 mer” seed binding site, a context score greater than 80 and both an “8 mer” site and a context score greater than 80. The comparison reveals significant overlap between predicted and identified sites in the datasets, especially when considering high-confidence TargetScan predictions. (c) Scatterplot of mouse \textsuperscript{CNV}K pulldown RNA-seq data, highlighting qPCR validated target genes. Each control gene is shown using two points - the circle demonstrates normalised counts detected using \textsuperscript{CNV}K-miR-29b, the triangle - \textsuperscript{CNV}K-scram. It is apparent that genes such as \textit{Dnmt3a} and \textit{Col1a1} demonstrate a strong shift towards the upper left quadrant, while non-target genes such as \textit{Gapdh} and \textit{Hprt} do not. A list of previously observed miR-29b targets was obtained by reviewing the literature, focusing on studies which had identified direct binding interactions and used luciferase assays to validate binding. (d) Scatterplot comparing the normalised number of reads for a given gene in the input (x-axis) and pulldown (y-axis) NIH3T3 datasets (mean of three replicate “lysed” experiments). Genes with a pulldown enrichment greater than four were considered to be targets of miR-29b in mouse (green filled circles). A subset of these were also found to be miR-29b targets in human cells, and are shown as black filled circles. It is apparent that likelihood of conservation is not dependent upon expression level.
Supplementary Figure 4: Quality control of sequencing datasets and comparison of mouse and human replicate datasets. Replicate experimental series are labelled 1-3 using the “lysed” protocol and 4 using the “intact” strategy. (a) Hierarchical clustering of mouse NIH3T3 CNV K generated datasets Heatmap showing Euclidean distances between the samples as calculated from the DESeq2 variance stabilised transformation based on the expression of the top 500 genes (considering the top 1000 or 10000 genes did not alter the clustering profile). It is apparent that the 29b pulldown datasets are most similar to each other - and distinct from all other libraries. (b) Comparison of intact and lysed crosslinking protocols - mouse NIH3T3 datasets Venn diagram of the overlap between the sets of genes characterised by a PE > 4 in each of the replicate experimental series. The four replicates display good concordance, and a comparable number of replicate-specific genes are observed across the three lysed and intact replicates. Three libraries were generated from HeLa cells, all of which utilised an “intact” protocol, where cross-linking occurred prior to cell lysis. (c) Hierarchical clustering of human HeLa CNV K generated datasets Heatmap showing the Euclidean distances between the samples as calculated from the DESeq2 variance stabilised transformation based on the expression of the top 500 genes. Using 1000 or 10000 genes for analysis resulted in a small increase between the scores observed for the six pulldown samples (however, scrambled pulldown remained more similar to input datasets than the 29b specific pulldown). (d) Overlap between the number of genes identified as miR-29b targets (PE > 4) in each of the three series of human experiments. It is
apparent that the majority of genes that are pulled down and identified as miR-29b targets are observed in all three datasets, with an additional 11-40% of hits unique to one or two of the datasets.
Supplementary Figure 5: Switchgear genomics luciferase assay validation 88 genes identified as miR-29b targets using the \textsuperscript{CNV} pulldown approach were interrogated with Switchgear LightSwitch luciferase assays, of which 64 were significantly down-regulated (FDR<0.05, names shown in green; Welch’s t-test with Benjamini-Hochberg multiple testing correction). This included 13/16 previously known control targets, highlighted in gray, as well as 51/72 novel targets. Genes were grouped based on ontology as described in Online Methods. Bar – median of three replicate experiments.
Supplementary Figure 6: Using the $^{\text{CNVK}}$ approach, many novel miR-29b targets are identified in canonical pathways regulated by this small RNA. Genes reported to be mouse miR-29b targets in the literature and observed using the $^{\text{CNVK}}$ pulldown approach are highlighted in red, previously identified targets not pulled down using $^{\text{CNVK}}$ are shown in blue, while novel hits identified using the $^{\text{CNVK}}$ approach are highlighted in green. (a) Mouse focal adhesion, (b) Human focal adhesion, (c) Human apoptosis pathway.
Supplementary figure 7: MicroRNA-29b is enriched in the nucleus of NIH3T3 mouse fibroblast cells. 
(a) Northern blot hybridization of NIH3T3 nuclear and cytoplasmic RNA samples probed for the presence of miR-29b, miR-15b, miR-16 and cytoplasmic marker tRNA Lys (AAG). Low molecular weight RNA marker (Affymetrix) was used for all Northern blot hybridizations. Dashed lines represent auto-contrast of relevant sections of the blot. 
(b) Bioanalyzer electropherograms obtained for total, nuclear and cytoplasmic RNA of NIH3T3 cells. RNA Integrity Number (RIN) scores are indicated. qRT-PCR analysis of nuclear, cytoplasmic RNA purity based on quantification of cytoplasmic marker tRNA Lys (AAG) and nuclear marker SNORA19. Bars represent mean ± SEM of three independent experiments. 
(c) Immunoblotting for cytoplasmic (Gapdh), nuclear (Histone H3, Lamin A/C) and endoplasmic reticulum membrane (Calnexin) proteins confirms efficient cellular fractionation. M refers to molecular weight marker (kDa) and arrows indicate expected size.
bands. (d) Splinted ligation demonstrates miR-15b, miR-16 and miR-29b expression in nuclear (Nuc.) and cytoplasmic (Cyto.) fraction of NIH3T3 cells. Cell equivalent amounts of nuclear and cytosolic RNA/protein were used for all analysis. T= total, N= nuclear, C= cytoplasmic RNA or lysate.
Choudhary et al.

**a**

Log2 nuclear read counts (CPMM) vs. Log2 cytoplasmic read counts (CPMM) for differentially expressed miRNAs.

**b**

Tag abundance distribution for different subcellular fractions.

**c**

Position of the 3′ end of the small RNA relative to the splice donor site (3′ end of the exon) (nt).

**d**

Flowchart of the experimental procedure:
- Transfect cells with CNVK containing miRNA mimics.
- miRNA-target pairs crosslinked.
- Hypotonic lysis.
- Wash nuclei.
- Collect cytoplasmic fraction.
- RNA quality check.
- RNA quantification.
- Bioanalyzer RNA Pico-chip analysis.
- RNA-Seq.

**e**

Bioanalyzer Gel Image showing cytoplasmic and nuclear small RNAs.

**f**

Heatmap showing log2 read counts for different subcellular fractions.

**g**

Venn diagram showing the overlap of subcellular fractions.

All intact and lysed 3T3 datasets:
- Cyto
- Nuc

Any intact or lysed 3T3 dataset:
- Cyto
- Nuc

Legend:
- CNVK
- K
- miRNA mimics
Supplementary Figure 8: RNA-seq analysis of CNVK-identified miR-29b targets in NIH3T3 nuclei. (a) RNA-Seq analysis of small RNA population in NIH3T3 nuclei and cytoplasm confirms miR-29b as one of the nuclear enriched miRNAs. Red dots are more than 2 fold enriched in either compartment. Blue dots are less than 2 fold enriched in either compartments and dots below 0 have very low counts in either libraries and are hence not considered for differential analysis. Inset is a zoomed-in view of the graph displaying miR-29b and other nuclear enriched miRNAs. (b) Small RNA size distribution and total abundance of unmapped tags after adaptor trimming. X-axis depicts small RNA tag size. Left Y-axis depicts total tag abundance; right Y-axis depicts number of distinct tags. (c) Splice-site RNA (spliRNA) abundance in NIH3T3 nuclear (top panel) and cytoplasmic (bottom panel) small RNA dataset. Position of small RNA 3′ ends is plotted with respect to the splice donor site (that is, the 3′ end of the exon). Diagrammatic representation of spliRNA position and strand orientation with respect to exon-exon junctions is indicated. (d) Illustration of CNVK oligo based nuclear and cytoplasmic microRNA target pulldown methodology. (e) Northern blot hybridization on subcellular fractions of NIH3T3 cells transfected with CNVK-miR-29b and CNVK-scram oligos. Blots were probed with miR-29b and scrambled probes respectively. Low molecular weight RNA marker (Affymetrix) was used for all blots. Entire yield of nuclear and cytoplasmic pulldown RNA was used for Northern blot hybridization. 1/10th amount of transfected cells (i.e. ~1 million cells) were lysed with RIPA buffer before isolation of pulldown RNA from whole cells, which is used as positive control (T). M= marker, T= total RNA, N= nuclear RNA, C= cytoplasmic RNA. (f) Hierarchical clustering of mouse nuclear and cytoplasmic CNVK datasets. Heatmap showing the Euclidean distances between the samples as calculated from the DESeq2 variance stabilised transformation based on the expression of the top 500 genes (using 1000 or 10000 did not significantly alter the clustering). (g) Comparison of cytoplasmic identified targets with high-confidence targets observed in all four NIH3T3 and at least in one of the NIH3T3 datasets. All 236 targets identified in the cytoplasmic fraction were tested for enrichment in the total NIH3T3 datasets, and a large proportion were identified as miR-29b targets utilising the two approaches. No targets were unique to the nucleus.
Supplementary Figure 9: Fold-change (FC) of known mouse miR-29b targets in CNVK datasets. This figure presents the observed enrichment of known miR-29b targets in CNVK datasets normalised to GAPDH expression levels, in the same format as was recently reported for a novel psoralen - based pulldown strategy\(^{25}\), Figure 2b). Actb is not a miR-29b target, while all other presented genes are. Most targets demonstrate a fold change between 5x and 25x in each of the replicate experiments. FC1, FC2, FC3 and FC4 are the four conducted mouse experimental series, 1 to 3 are lysed; 4 is intact. Boxplots demonstrate the overall FC distribution. The Imig et al. (2015) study was characterised by a median of 4x and maximum of 9x enrichment, and it is apparent that CNVK pulldowns result in a much stronger fold enrichment.
Supplementary Figure 10: Crosslinked structure of photoactivable nucleosides to aromatic amino acid side chain. (a) UV 365nm induced 4-thiouridine (4SU) - aromatic amino acid crosslinking. (b) Predicted structure of UV 365nm induced $\text{CNV}_K$-aromatic amino acid crosslinked product. Crosslinked structures for Tyrosine are shown. Tyr, Phe, Trp refer to aromatic amino acids tyrosine, phenlyalanine and tryptophan respectively.
Supplementary Figure 11: Immunoblot digital images. Full-length images of immunoblots (related to Figure 1d and Supplementary Figure 2b) with the molecular weight ladders are shown. Boxes highlight the lanes used in the figures.
Supplementary Figure 12: Schematic diagram indicating primer binding sites for NIH3T3 qRT-PCR positive and negative control genes. Schematic diagram of the mRNAs of pre-validated mmu-miR-29b target genes used as positive controls for qRT-PCR: Col1a1, Col3a1, Col5a3, Dnmt3a and Adam12. Negative control genes were: Gapdh, Hprt1 and Irf6. The untranslated regions of the mRNA are indicated in green, while the translated regions are indicated in pink. Exon numbers are indicated below each cylinder, first exon is the 5'UTR and last indicates the 3'UTR, where miR-29b binding sites are present. Grey line depicts intronic region, arrows indicate forward and reverse primers and asterisk *** indicates TargetScan predicted miR-29b binding sites in 3'UTR region. Transcript information is obtained from UCSC genes, UCSC Genome Browser Mouse mm10 assembly.