Identification of miR-29b targets using 3-cyanovinylcarbazole containing mimics

ANUPMA CHOUDHARY,[1,2,7] DARYA P. VANICHKINA,[1,3,7] CHRISTINE ENDER,[1] JOANNA CRAWFORD,[1] GREGORY J. BAILLIE,[1] ANDREW D. CALCINO,[4] KELIN RU,[1] and RYAN J. TAFT[1,5,6]

[1]Institute for Molecular Bioscience, The University of Queensland, Brisbane, Queensland 4072, Australia
[2]CSIRO Health and Biosecurity, Australian Animal Health Laboratory, Newcomb 3220, Australia
[3]Gene and Stem Cell Therapy Program, Centenary Institute, University of Sydney, Sydney 2050, Australia
[4]Department of Integrative Zoology, University of Vienna, Vienna 1090, Austria
[5]School of Medicine and Health Services, Departments of Integrated Systems Biology and of Pediatrics, George Washington University, Washington DC 20052, USA
[6]Illumina, Inc., San Diego, California 92122, USA

ABSTRACT

MicroRNAs (miRNAs) are highly conserved ~22 nt small noncoding RNAs that bind partially complementary sequences in target transcripts. MicroRNAs regulate both translation and transcript stability, and play important roles in development, cellular homeostasis, and disease. There are limited approaches available to agnostically identify microRNA targets transcriptome-wide, and methods using miRNA mimics, which in principle identify direct miRNA:transcript pairs, have low sensitivity and specificity. Here, we describe a novel method to identify microRNA targets using miR-29b mimics containing 3-cyanovinylcarbazole (^{CNV}K), a photolabile nucleoside analog. We demonstrate that biotin-tagged, ^{CNV}K-containing miR-29b (^{CNV}K-miR-29b) mimics are nontoxic in cell culture, associate with endogenous mammalian Argonaute2, are sensitive for known targets and recapitulate endogenous transcript destabilization. Partnering ^{CNV}K-miR-29b with ultra-low-input RNA sequencing, we recover ~40% of known miR-29b targets and find conservation of the focal adhesion and apoptotic target pathways in mouse and human. We also identify hundreds of novel targets, including *NRAS*, *HOXA10*, and *KLF11*, with a validation rate of 71% for a subset of 73 novel target transcripts interrogated using a high-throughput luciferase assay. Consistent with previous reports, we show that both endogenous miR-29b and ^{CNV}K-miR-29b are trafficked to the nucleus, but find no evidence of nuclear-specific miR-29b transcript binding. This may indicate that miR-29b nuclear sequestration is a regulatory mechanism in itself. We suggest that ^{CNV}K-containing small RNA mimics may find applicability in other experimental models.

Keywords: 3-cyanovinylcarbazole; ^{CNV}K; miR-29b; mimics; target

INTRODUCTION

3-Cyanovinylcarbazole (^{CNV}K) is a recently described nucleoside analog that rapidly photocrosslinks to complementary pyrimidines upon exposure to 365 nm UV light (Supplemental Fig. 1a), with a reported in vitro crosslinking efficiency approaching 100% (Yoshimura and Fujimoto 2008). Recent work has demonstrated ^{CNV}K utility in a range of biological applications, including plasmid labeling (Fujimoto et al. 2012) and transient transgene silencing (Sakamoto et al. 2014), but its use as a mechanism to identify targets of endogenous small RNAs has not yet been explored.

microRNA-29b (miR-29b) is one of the most highly conserved and multipotent microRNAs identified. It has been linked to a wide range of biological processes—spanning from stem cell differentiation to extracellular matrix remodeling and apoptosis (Li et al. 2009b, 2013; Suh et al. 2012)—and its sequence is invariant across Bilateria (Supplemental Fig. 1b). miR-29b targets a diverse set of genes including transcription factors, collagens, and methyltransferases, and its dysfunction is associated with matrix-associated disorders, oncogenesis, and metastasis (Park et al. 2009; Zhang et al. 2014). miR-29b has also been shown to localize to the animal nucleus (Hwang et al. 2007), but its function in this cellular compartment has remained unclear.

RESULTS

We sought to use ^{CNV}K to identify miR-29b targets transcriptome-wide and generated single stranded 3′ biotin-tagged and ^{CNV}K-containing mature miR-29b mimics (Fig. 1A). Recent studies have reported the use of ^{CNV}K-containing oligonucleotides in live cells (Sakamoto et al. 2014), but a systematic assessment on cell viability has not been performed. We transfected NIH3T3 cells with increasing concentrations of ^{CNV}K-miR-29b duplex (Supplemental Fig. 1c) and found no significant differences in viability for cells transfected with up to 40 nM concentration of ^{CNV}K-miR-29b at 24 (Fig. 1B) or 48 h post-transfection (Supplemental Fig. 1d). To assess if ^{CNV}K itself, or the sequence composition of the oligonucleotide, affected cellular uptake, we performed northern blot analysis of cells transfected with equimolar amounts of ^{CNV}K-miR-29b or a 21 nt biotin-tagged ^{CNV}K-containing control oligonucleotide (^{CNV}K-scram), and found that both were equally internalized (Supplemental Fig. 1e). To further test the impact of ^{CNV}K-miR-29b on endogenous miRNA levels, cells were transfected with increasing concentrations of ^{CNV}K-miR-29b or ^{CNV}K-scram (Fig. 1C; Supplemental Fig. 2a). We observed a dose-dependent increase in intracellular miR-29b concentration only in cells that were transfected with ^{CNV}K-miR-29b, and no change in endogenous miR-16 levels was observed in the presence of either ^{CNV}K-miR-29b or ^{CNV}K-scram.

MicroRNAs act as the RNA guide for Argonaute2, the central component of the RNA induced silencing complex, which in the case of miR-29b may suppress translation or destabilize target transcripts (Chou et al. 2013). In order to investigate if ^{CNV}K-miR-29b binds endogenous Argonaute2, we performed pulldown assays in HeLa and NIH3T3 cells transfected with equimolar amounts of ^{CNV}K-oligonucleotide. We observed a specific association of ^{CNV}K-miR-29b with Argonaute2 in both human and mouse (Fig. 1D; Supplemental Fig. 2b). To examine if ^{CNV}K-miR-29b repressed endogenous mRNA levels of target genes we transfected cells with ^{CNV}K-miR-29b, ^{CNV}K-scram, a commercially available native miR-29b duplex (n-miR-29b), or a nontargeting control (*C. elegans* miR-67). Consistent with the reports of other miR-29b targets (Chou et al. 2013), the expression level of known miR-29b target transcripts *Col3a1*, *Col1a1*, *Col5a3*, *Dnmt3a*, and *Adam12* were significantly reduced after transfection with ^{CNV}K-miR-29b or n-miR-29b, and not with negative control oligonucleotides ^{CNV}K-scram or *C. elegans* miR-67 (Fig. 1E; Supplemental Fig. 2c,d).

To assess if ^{CNV}K crosslinking can be used to enrich for target transcripts we performed biotin pulldown assays using ^{CNV}K-miR-29b, ^{CNV}K-scram and biotinylated-miR-29b (bio-miR-29b) in both intact and lysed cells after UV exposure. Compared to ^{CNV}K-scram and biotinylated-miR-29b (bio-miR-29b) we observed a 3.9- to 21-fold increased enrichment of known target transcripts with ^{CNV}K-miR-29b after 10 min or more of UV irradiation (Fig. 1F,G;

Supplemental Fig. 2e). Surprisingly, ^{CNV}K-miR-29b samples that were not exposed to UV also displayed a moderate enrichment of target genes compared to controls (Supplemental Fig. 2f), suggesting that ^{CNV}K crosslinking may occur under ambient light conditions. Together these observations indicate that ^{CNV}K-miR-29b mimics can effectively recapitulate endogenous activity and enable identification of target transcripts in situ.

To identify genome-wide targets of miR-29b we partnered ^{CNV}K-miR-29b pulldowns with low input RNA-sequencing. In brief, ^{CNV}K-oligonucleotides were transfected into mammalian cells and irradiated for 10 min, and isolated RNA was prepared for sequencing using the Epicentre TotalScript RNA Kit with 5 ng input using a 1:1 mixture of random hexamer and oligo(dT) primers (Supplemental Note 1). For each ^{CNV}K-miR-29b target enrichment experiment a total of four samples were collected: ^{CNV}K-miR-29b and ^{CNV}K-scram target pulldowns and the corresponding input samples for each. Analysis of normalized gene-level read counts revealed that in intact HeLa and NIH3T3 cells a subset of transcripts segregated in the ^{CNV}K-miR-29b pulldown samples when compared to either input or ^{CNV}K-scram libraries (Fig. 2A,B; Supplemental Fig. 3c,d).

To robustly identify ^{CNV}K-miR-29b associated transcripts we calculated a pulldown enrichment (PE) score for each transcript of interest, which assesses relative enrichment of a transcript in the ^{CNV}K-miR-29b pulldown against the background of the input and relative to ^{CNV}K-scram (Fig. 2A,B; Supplemental Fig. 4, and Online Methods). We used a cutoff of PE >4, which ensured that a putative miR-29b target was at least fourfold more highly represented in the ^{CNV}K-miR-29b pulldown compared to ^{CNV}K-scram, even when the ^{CNV}K-miR-29b read counts were corrected for input transcription levels. This resulted in a set of 567 mouse and 827 human putative miR-29b target transcripts (Supplemental Table 1). The vast majority of enriched transcripts were protein-coding ($P < 0.0001$, χ^2; Supplemental Table 2), although the p53-regulated *TUG1* lincRNA was among the ncRNAs identified as a putative miR-29b target in both species. Naïve microRNA target enrichment analysis of these transcripts using ToppGene (Chen et al. 2009) revealed miR-29b as the most significant upstream regulator (216/522 in mouse and 316/763 genes in human; Supplemental Table 3). These target genes were also significantly overrepresented in a list of TargetScan miR-29b predictions (permutation test, P-value = 0.00009999) (Supplemental Fig. 3a,b; Supplemental Table 7), displayed a significant overrepresentation of sequences complementary to miR-29b seed region (Supplemental Note 2) and, included >40% of expressed miR-29b targets previously reported in the literature (Supplemental Table 4).

We selected 65 novel ^{CNV}K-miR-29b enriched human targets and 23 known positive controls across the complete range of expression levels, conservation scores, and TargetScan-binding site types for validation by Switchgear Lightswitch luciferase assay. All seven negative controls
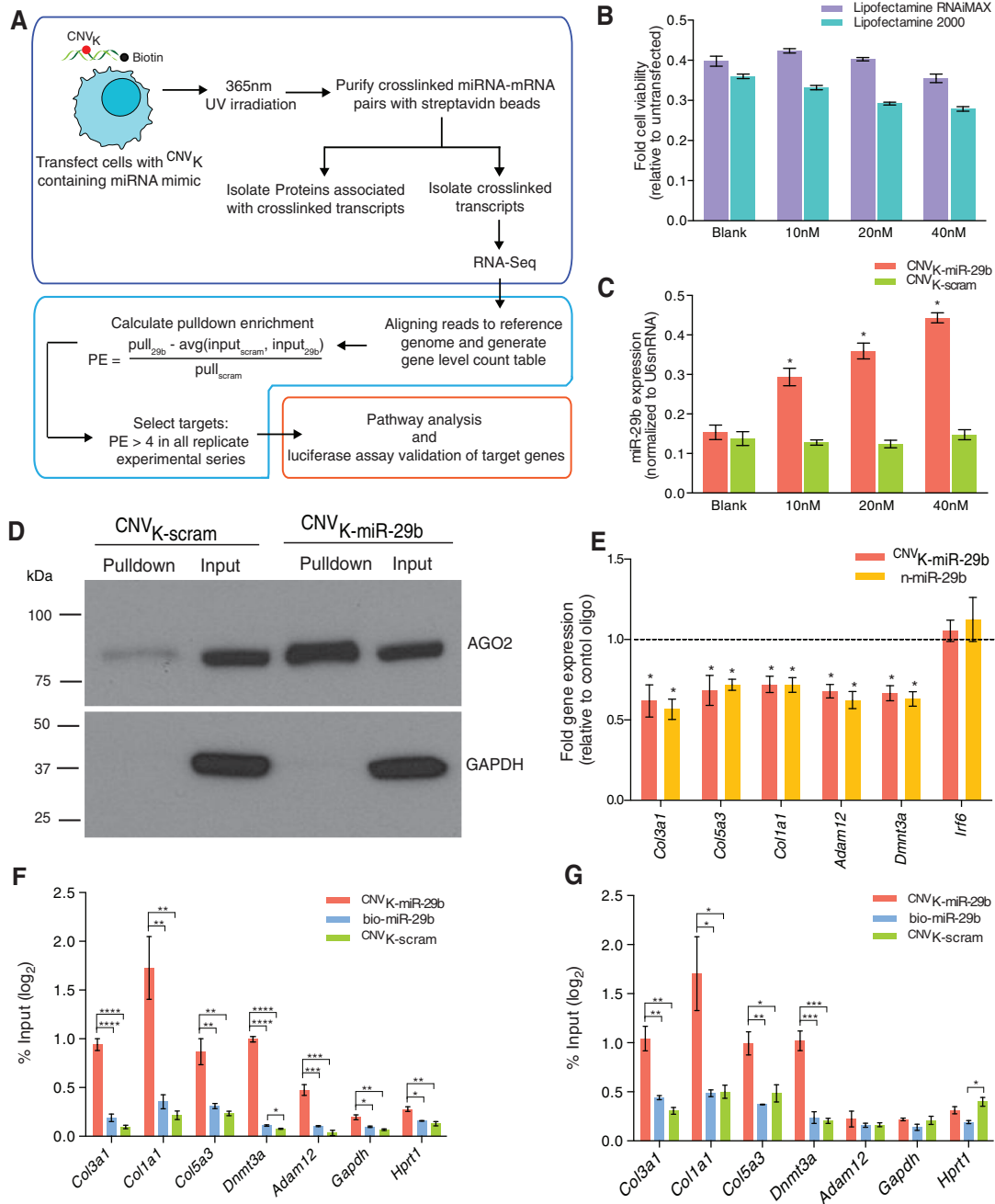
FIGURE 1. 3-Cyanovinylcarbazole-based miRNA target pulldown. (*A*) ^{CNV}K mimic-based microRNA target pulldown methodology. (*B*) MTT assay-based assessment of NIH3T3 cell viability measured 24 h after transfection with ^{CNV}K-miR-29b. (*C*) TaqMan qRT-PCR based assessment of miR-29b expression (normalized to *U6* expression) in untransfected cells and cells transfected with increasing doses of ^{CNV}K-miR-29b and ^{CNV}K-scram. (*) $P < 0.05$ compared to ^{CNV}K-scram; two-tailed Student's *t*-test. (*D*) Representative immunoblot demonstrating association of ^{CNV}K-miR-29b and ^{CNV}K-scram with Argonaute2 protein in HeLa cells. Of note, 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 nonspecific protein pulldown. The antibody used for HeLa lysate is anti-human α-AGO2 antibody (ab57113, Abcam). M refers to molecular weight marker (kDa). (*E*) NIH3T3 cells were transfected with 20 nM concentration of ^{CNV}K-miR-29b, ^{CNV}K-scram (negative control oligo), n-miR-29b, or *C. elegans* miR-67 (negative miRNA control) oligos, and gene expression values were normalized to *Hprt1*. Gene expression levels in ^{CNV}K-miR-29b transfected cells are plotted as a ratio of respective gene levels in ^{CNV}K-scram transfected cells (set at one), 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 one). *Irf6* is used as nontarget control gene. (*) $P < 0.05$ compared to respective negative control oligos; two-tailed Student's *t*-test. qRT-PCR analysis of previously reported miR-29b target gene expression in NIH3T3 cells transfected with 10 nM ^{CNV}K-miR-29b, bio-miR-29b, or ^{CNV}K-scram oligonucleotides. Transfected cell lysates were irradiated with 365 nm UV light for either (*F*) 10 min or (*G*) 20 min to induce covalent crosslinking of ^{CNV}K with complementary nucleotides in target mRNA. (*) $P < 0.05$, (**) $P < 0.01$, (***) $P < 0.001$, (****) $P < 0.0001$ compared to other transfected cells; one-way ANOVA, followed by Tukey's post hoc test. Bars represent mean ± SEM of three independent experiments.
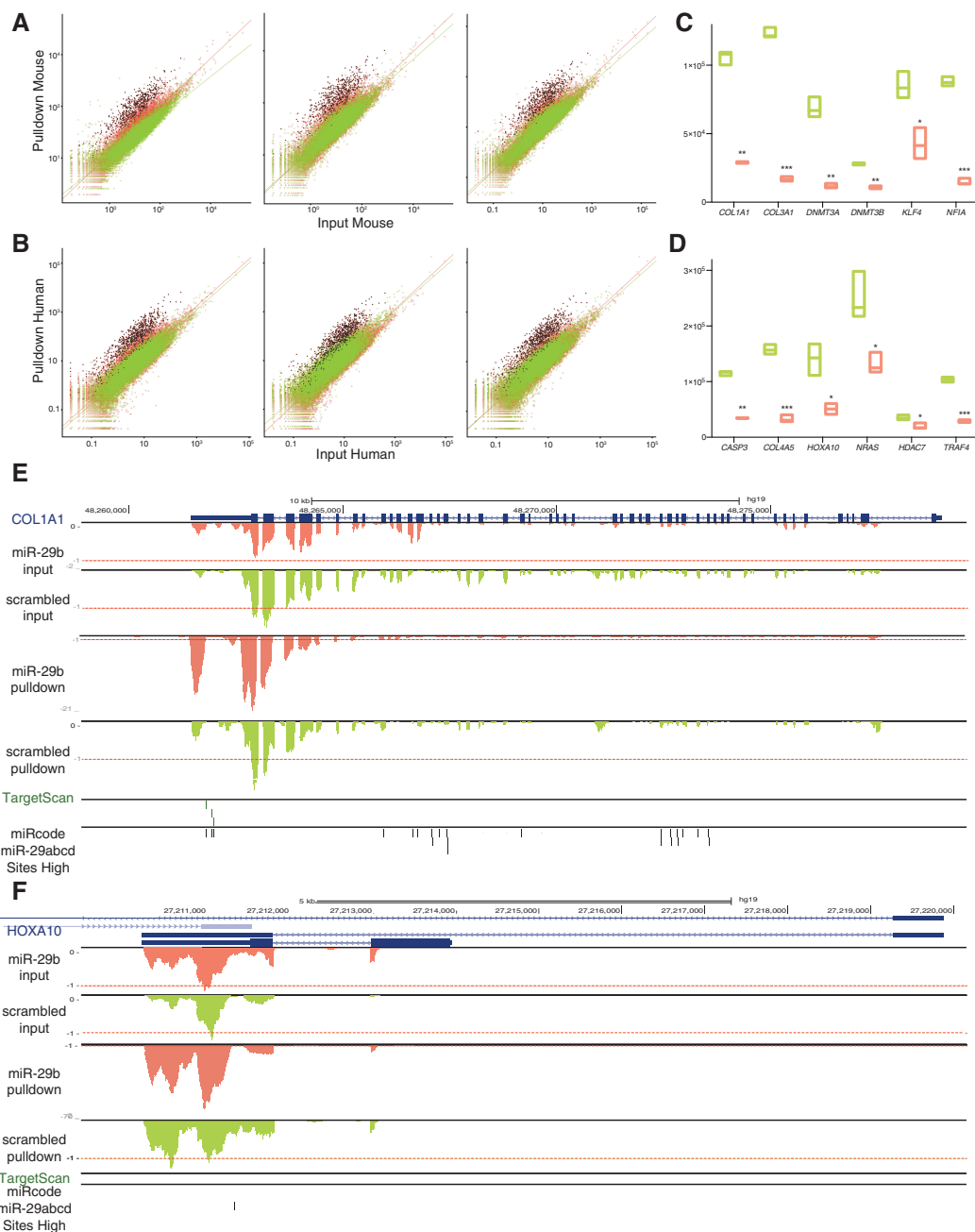
FIGURE 2. Discovery of novel miRNA targets using 3-cyanovinylcarbazole. (*A*) Scatterplots of three mouse replicate ^{CNV}K pulldown RNA-seq experimental series, highlighting the 567 genes identified as mouse miR-29b targets in black. Each scatterplot shows the number of reads for a given gene in the input (*x*-axis) and pulldown (*y*-axis), for the ^{CNV}K-miR-29b (red) and ^{CNV}K-scram (green) samples, respectively. (*B*) Scatterplots of three human replicate ^{CNV}K pulldown RNA-seq experimental series, highlighting the 827 genes identified as human miR-29b targets in black. Axes/color as in panel *A*. (*C*) Validation of previously known miR-29b targets observed in the ^{CNV}K pulldown using Switchgear LightSwitch luciferase assays. A total of 88 genes were interrogated using this assay, of which 64 were found to be significantly down-regulated with a FDR < 0.05 after Benjamini–Hochberg multiple testing correction. For each of the six previously known target genes shown in this panel, the normalized luciferase intensity is presented for the miR-29b mimic (red) and nontemplate control oligonucleotide (green). Median shown as bar; (*) $P_{adj} < 0.05$, (**) $P_{adj} < 0.01$, (***) $P_{adj} < 0.001$, (****) $P_{adj} < 0.0001$. (*D*) Validation of novel miR-29b targets discovered using the ^{CNV}K pulldown using Switchgear LightSwitch luciferase assays. Of note, 52/73 novel genes investigated were validated using the luciferase assay with an FDR < 0.05 (Supplemental Fig. 5; Supplemental Table 5), with six representative examples shown here. (*E*) UCSC Genome Browser snapshot of read density, TargetScan and miRcode predictions in the human *COL1A1* locus, a previously known miR-29b target gene, and for (*F*) *HOXA10*, a novel miR-29b target identified with the ^{CNV}K approach. Wiggle tracks normalized to 100 million reads per library reveal read density across the locus in four libraries from one of the replicate experimental series: two input data sets (miR-29b and scrambled input) and pulldown data sets with the miR-29b mimic and scrambled oligonucleotide are shown, and TargetScan and miRcode miR-29b prediction tracks are displayed *below*. The red dashed line corresponds to a normalized peak height of one in all four of the experimental panels, revealing the enrichment of reads mapping to the *COL1A1* and *HOXA10* loci specifically in the miR-29b pulldown, and not in the input data sets or scrambled pulldown.

showed no change in luciferase activity, while 64 out of 88 (73%) ^{CNV}K–miR-29b targets were validated (Fig. 2C–E; Supplemental Fig. 5, Supplemental Table 5)—including 12/15 (80%) previously reported targets (Supplemental Fig. 9) and 27/35 targets (77%) conserved between mouse and human (Supplemental Table 1). Targets with higher expression levels or PE values were not more likely to be functionally validated (Pearson's χ^2 test *P*-value = 0.9983), indicating that ^{CNV}K–miR-29b pulldowns are able to enrich for lowly expressed target RNAs. No correlation was observed between predicted binding site type and validation likelihood (Pearson's χ^2 test *P*-value = 0.9965), supporting recent studies reporting a high prevalence of noncanonical miRNA-binding interactions (Hafner et al. 2010; Loeb et al. 2012; Helwak et al. 2013). Indeed, 24 of 33 (72%) ^{CNV}K–miR-29b targets for which the Switchgear 3′ UTR construct did not contain a TargetScan predicted binding sites were validated, as well as 40/55 (73%) of those that did—indicating that ^{CNV}K–miR-29b can successfully identify targets with both canonical and noncanonical-binding sites. These data demonstrate that ^{CNV}K–miR-29b is capable of identifying bona fide miR-29b target transcripts with high confidence.

miR-29b has been shown to be a key regulator of genes involved in cell motility, focal adhesion, apoptosis modulation and cancer (Kriegel et al. 2012; Wang et al. 2014), and analysis of both human and mouse ^{CNV}K–miR-29b target transcripts reflected these reports. Mouse targets were over-represented in integrin signaling, assembly of collagen fibrils, and focal adhesion. Human targets were enriched in genes involved in apoptosis modulation and signaling, DNA damage response, and the p53 signaling pathway (Supplemental Fig. 6; Supplemental Table 3). It is likely that the tissue of origin of the cell lines used (fibroblast-derived mouse NIH3T3 and cervical cancer derived HeLa) rather than species-specific miR-29b targeting, underpin the observed differences in ontology enrichment. We were able to identify dozens of novel targets (shown in green in Supplemental Fig. 6) that belong to these or other pathways. For example, an additional 15 members of the human apoptosis pathway were identified as miR-29b targets, a greater than threefold increase in the number of targets in this pathway. Likewise, the repertoire of collagens targeted by miR-29b was expanded to include *COL15A1*, *COL4A5*, and *COL5A1*, with both known and novel collagens validated by luciferase assay as described above. This suggests that the majority of collagens, and perhaps many genes involved in extracellular matrix reorganization (Supplemental Table 3) are miR-29b targets. We also identified targets in developmental and transcription factor gene families including five Kelch-like (KLHL) family, four Kruppel-like factor (KLF), eight zinc finger with C2H2 and BTB domain (ZBTB), three Frizzled (FZD), and 19 ring type zinc finger (RNF) genes in the human and mouse ^{CNV}K–miR-29b data sets (Supplemental Table 3)—indicating that miR-29b is involved in the regulation of a wide range

of fundamental biological processes. Select additional novel miR-29b targets include *NREP*, an RNA-binding protein that forms a complex with the 5′ UTR of *TGF-β* and *eIF3b* to stimulate the translation of *TGF-β1*, 2, and 3 (Yue et al. 2014) and can induce epithelial–mesenchymal transition in various human cancers (McDonough et al. 2005; Peng et al. 2012). Likewise, *HoxA10* (Fig. 2D,F), a master regulator of postnatal hematopoietic development (Magnusson et al. 2007) is highly enriched in the ^{CNV}K–miR-29b data set, but does not contain a predicted miR-29b-binding site.

To investigate the potential nuclear role of miR-29b, the nuclear localization of miR-29b in NIH3T3 cells was investigated using three independent approaches—northern blot, splinted ligation, and small RNA sequencing (Supplemental Figs. 7, 8a–c). Northern blotting confirmed that transfected ^{CNV}K–miR-29b, but not ^{CNV}K-scram, localized to the nucleus (Supplemental Fig. 8e). We performed in situ nuclear UV crosslinking of ^{CNV}K–miR-29b in NIH3T3 cells and completed the pulldown, sequencing and analysis as described above (Supplemental Fig. 8d,f,g). We failed to identify any well-supported nuclear-specific transcripts with PE scores greater than four, and those close to this threshold were found at high enrichment in the cytoplasmic data sets (Supplemental Table 1), which may suggest that miR-29b does not have specific RNA targets in the nucleus.

DISCUSSION

Here we have described a novel approach for the identification of miR-29b targets using ^{CNV}K-oligonucleotides. Coupling this approach with ultra-low input RNA sequencing enabled successful identification of miR-29b targets in mouse and human cell lines, which has extended the miR-29b targetome by uncovering dozens of new targets belonging to key developmental and regulatory gene families. Our data also suggest that microRNA localization to the nucleus is not related to specific transcript targets. We propose that nuclear localization may be a mechanism to sequester these species from their cytosolic targets—a hypothesis that is particularly appealing for miR-29b since quick release from the nucleus could enable rapid matrix reorganization, and this could explain the pronounced nuclear localization of miR-29b in fibroblast cells. It is also possible that transient nuclear sequestration is a general feature of miRNA regulation, which could explain the increasing number of studies showing that miRNAs (and components of the RISC complex) are frequently observed in the nucleus (Jeffries et al. 2010; Liao et al. 2010; Gagnon et al. 2014).

^{CNV}K may crosslink with proteins (Supplemental Fig. 10) and could be coupled with HITS–CLIP or PAR–CLIP to identify RNA-protein interacting complexes, or partnered with iCLIP or CLASH for precise identification of miRNA binding sites (Licatalosi et al. 2008; Hafner et al. 2010; Konig et al. 2010; Helwak et al. 2013). Since the ^{CNV}K-

approach enables purification of full-length target RNAs, it can be utilized to increase signal-to-noise ratios; for example, ^{CNV}K-containing mimics could be used to stabilize probe-target interactions in RNA–FISH experiments. It should be noted that ^{CNV}K has only been reported to crosslink with pyrimidines (Yoshimura and Fujimoto 2008) and potentially limits the application of this protocol to microRNAs with A or G at position 8. Future studies should systemically examine the effect of ^{CNV}K placement within the synthetic miRNA oligo, the impact of sequence content and 3′ target base-pairing on both ^{CNV}K crosslinking and target recovery efficiency and the relationship between seed-type and ^{CNV}K efficiency. Overall, we anticipate that incorporation of ^{CNV}K and other similar high-efficiency nucleoside analogs into RNA targeting experiments will yield additional insights into RNA biology.

MATERIALS AND METHODS

Oligonucleotide synthesis and annealing

n-miR-29b and *C. elegans* miR-67 were obtained as annealed duplexes from Dharmacon (GE Healthcare), while bio-miR-29b (Integrated DNA Technologies) and ^{CNV}K containing oligonucleotides (Nihon Techno Service) were obtained as HPLC purified dried RNA pellets of sense and antisense strands (see Supplemental Table 6 for oligonucleotide sequences). All strands contained 5′ terminal phosphate and where present, biotin was attached to the 3′ OH group of the sense strand via a TEG (tetra-ethylene-glycol, 15 atom) spacer arm. Sense and antisense strands were annealed by mixing equimolar amounts of each oligonucleotide with oligo annealing buffer (20 mM potassium acetate, 6 mM HEPES pH 7.4, 400 μM magnesium acetate). This resultant solution was incubated in boiling water that was left to cool down slowly at room temperature. Annealed duplexes were diluted before storing as aliquots. Successful annealing of oligos was confirmed by running annealed duplexes alongside single stranded sense and antisense oligos on a 2% agarose gel.

MTT assay

Exponentially growing NIH3T3 cells (7000 cells/well) were reverse transfected in 96-well plates with varying doses of miR-29b ^{CNV}K oligos using either Lipofectamine 2000 (Life Technologies) or Lipofectamine RNAiMAX (Life Technologies). After 24 or 48 h post-transfection, cells were treated with 1 mg/mL MTT (Sigma) for 4 h at 37°C. Following removal of MTT reagent, cells were treated with 100% isopropanol for 40 min and the spectroscopic absorbance of colored extract was measured at 570 nm. Empty wells with medium only served as blank. Cells treated with growth medium only were called "Untreated" and represented healthy growing cells. "Reagent only" cells were treated with either of the two lipid-based transfection reagents and were used to assess the reduction in cell viability caused by transfection reagents alone in the absence of any ^{CNV}K oligonucleotide. The fold cell viability was calculated relative to the untreated control cells.

Cell growth and transfection

HeLa and NIH3T3 cells were maintained in DMEM media supplemented with 10% (v/v) FBS and 2 mM L-glutamine. All cells were obtained from American Type Culture Collection and cultured at 37°C and 5% (v/v) CO_2. All cells were tested negative for mycoplasma contamination (MycoAlert Mycoplasma Detection Kit, Lonza). One million cells were reverse transfected with equimolar amounts of ^{CNV}K-miR-29b and ^{CNV}K-scram oligos (Nihon Techno Service) per well in a 6-well plate using Lipofectamine RNAiMax reagent (Life Technologies) according to the manufacturer's instructions. Unless mentioned otherwise, mimics were transfected at a final concentration of 10 nM. Cells were topped up with 2 mL of complete growth media 24 h after transfection, and were allowed to recover for an additional 8 h before harvesting. To prepare deep sequencing libraries from pulldown RNA, approximately 9×10^6 cells were transfected with ^{CNV}K mimics. For qRT-PCR analysis of target genes, 6×10^6 cells were transfected, while 4×10^6 cells were transfected for analysis of ^{CNV}K oligo associated proteins by immunoblotting.

RNA and protein pulldown using ^{CNV}K oligos

NIH3T3, HeLa cells were transfected with 10 nM annealed oligonucleotides in 6-well format using Lipofectamine RNAiMax reagent (Life Technologies). Twenty-four hours post-transfection, cells were topped up with 2 mL complete growth media and allowed to recover for 8 h. For "lysed" crosslinking protocol, cells were rinsed in ice-cold PBS twice, harvested by scraping, and lysed in hypotonic lysis buffer (10 mM KCl, 1.5 mM $MgCl_2$, 10 mM Tris-Cl pH 7.5, 5 mM DTT, 0.5% NP-40, 60 U/mL SUPERase•In [Ambion] and 1× Complete Mini protease inhibitor cocktail [Roche]) (Cloonan et al. 2011). Cell debris was cleared by centrifugation (12,000*g* at 4°C for 2 min). The cleared lysate was irradiated as droplets on an ice-cold aluminum block for 10 min under a 20 W, 365 nm UV lamp at a distance of ∼10 cm from the light source, and 1/20th of its volume was kept as Input. The remaining irradiated supernatant was transferred to a clean tube to which NaCl was added to a final concentration of 1 M. Dynabeads MyOne Streptavidin C1 (12 μL beads/ million cells; Life Technologies) were preblocked with 1 μg/μL bovine serum albumin (Sigma) and 1 μg/μL yeast tRNA (Sigma), and incubated with the supernatant for 30 min at room temperature. Beads were then washed with hypotonic lysis buffer supplemented with 1 M NaCl before extracting associated RNAs from pulldown and Input samples using TRIzol LS (Life Technologies) (supplemented with 17 μg/mL Glycoblue [Ambion] and 10 mM $MgCl_2$), as per manufacturer's instructions.

For "intact" crosslinking protocol, cells were rinsed with ice-cold PBS once, 500 μL PBS was added per well, and living cells were irradiated on ice as above. Crosslinked cells were lysed in hypotonic lysis buffer and cell debris was cleared by centrifugation, as described above. One-twentieth of the volume of the supernatant was kept as Input to which TRIzol LS (Life Technologies) (supplemented with 17 μg/mL Glycoblue [Ambion] and 10 mM $MgCl_2$) was added, and RNA extraction was carried out as per manufacturer's instructions. The remaining supernatant was treated for biotin pulldown using Dynabeads MyOne Streptavidin C1 followed by RNA extraction as described above. RNA extraction, reverse transcription, and qRT-PCR were performed as described below, and one-way ANOVA

followed by Tukey's post-hoc test for multiple testing correction was used to test statistical significance among the three groups of cells transfected with ^{CNV}K-miR-29b, n-miR-29b, and ^{CNV}K-scram oligos, respectively. For statistical analysis, data were logarithmically transformed to ensure log-normal distribution, and a Brown-Forsythe test was used to confirm homogeneity of variance.

For isolation of proteins associated with ^{CNV}K oligos, cell transfection, crosslinking, and pulldown were performed as described above. For eluting proteins associated with ^{CNV}K oligos from pulldown and input samples, the magnetic beads were treated twice with 0.1% SDS and incubated at 95°C/5 min before collecting the eluate. The collected eluate was mixed with 4× Laemmli sample buffer (Bio-Rad) before immunoblotting.

Nuclear cytoplasmic fractionation

Cytoplasmic and nuclear RNA were isolated from NIH3T3 cells as described previously (Hwang et al. 2007). Briefly, the cells were trypsinized and harvested by centrifugation. The cell pellet was washed twice with ice cold PBS before incubating in buffer A (10 mM Tris-Cl pH 8, 140 mM NaCl, 1.5 mM MgCl$_2$, 0.5% NP-40) on ice for 5 min. After centrifugation at 1000g for 3 min, the cytoplasmic fraction was contained in the supernatant while the nuclear fraction was in the pellet. The cytoplasmic fraction was collected and spun again to remove any contaminating nuclear pellet. The nuclear pellet was washed twice in buffer A, and finally with buffer A containing 1% Tween-40 and 0.5% deoxycholic acid.

RNA from cytoplasmic supernatant and nuclear pellet was isolated using TRIzol LS, as per manufacturer's instructions. RNA quality and yield was assessed on an Agilent 2100 Bioanalyzer using the RNA 6000 Nano Chip Kit (Agilent). Cell equivalent amounts of RNA from nuclear and cytoplasmic fractions were used for all analyses, unless stated otherwise. RIPA lysis buffer (50 mM Tris-Cl pH 7.4, 150 mM NaCl, 1 mM EDTA, 1% Triton-X 100, 1% deoxycholic acid, 0.1% SDS) containing 1× cOmplete Mini Protease Inhibitor Cocktail (Roche) and 1× Phosphatase Inhibitor (1 mM sodium vanadate, 1 mM sodium pyrophosphate, 1 mM sodium molybdate, 10 mM sodium fluoride) was used to extract proteins from the nuclear pellet. Protein concentrations in cytoplasmic supernatant and nuclear lysate were determined using the BCA Protein Assay Kit (Pierce), according to the manufacturer's protocol. Ten micrograms of protein from nuclear and cytoplasmic lysate was used for immunoblotting.

RNA extraction, reverse transcription, and qRT-PCR

Total RNA was extracted from transfected cells using the mirVana miRNA Extraction Kit (Life Technologies). Concentrations of RNA samples were measured at 260 nm using a ND1000 Nanodrop (Biolab). RNA samples were DNase treated using TURBO DNase (Ambion) and purified using the RNeasy MinElute Cleanup Kit (Qiagen), to retain small RNAs according to manufacturer's instructions. cDNA was synthesized from 500 ng of extracted RNA using random hexamers at a concentration of 250 ng per 5 μg of RNA with the SuperScript III Reverse Transcriptase Kit (Invitrogen) as per manufacturer's instructions. The synthesized cDNA was diluted four times in RNase-free water and subsequently used for qRT-PCR to quantify mouse miR-29b target genes *Col3a1*, *Col1a1*, *Col5a3*, *Adam12*, and *Dnmt3a* (see

Supplemental Table 6 for primer sequences). Supplemental Figure 12 demonstrates primer binding sites for NIH3T3 qRT-PCR positive and negative control genes. Transcript levels for specific genes were quantified relative to expression of the housekeeping gene *Gapdh*, and data were expressed relative to *Gapdh* using the ΔCt method. qRT-PCR was carried out using the SYBR Green PCR Master Mix (Applied Biosystems) in 20 μL reactions, using 2 μL of diluted cDNA, 4 μL of primers (2 μM), 4 μL water, and 10 μL 2× master mix per reaction. For mature miRNA quantification, 5 ng of RNA was used for cDNA synthesis. Reverse transcription and qRT-PCR was carried out using the Taqman Small RNA Assays Kit (Life Technologies) as per the instructions of the manufacturer. MicroRNA expression was calculated relative to *U6* snRNA expression. All reactions were carried out in triplicates in 96-well plates on the ViiA7 system (Applied Biosystems).

The enrichment of tRNA-Lysine in the cytosol and snoRA19 in the nucleus was measured by qPCR to validate the purity of nuclear and cytoplasmic fractions. One microgram of Total RNA and cell equivalent amounts of nuclear and cytoplasmic RNA were treated with DNase I (Invitrogen), and 500 ng of it was primed using random hexamers at a concentration of 250 ng per 5 μg of RNA. cDNA was prepared using the SuperScript III Reverse Transcriptase Kit (Invitrogen) as per manufacturer's instructions. Due to the lack of an invariant internal control, we performed absolute quantification for the above RNAs. Briefly, the qPCR targets were PCR amplified for 35 cycles and purified from a 2% agarose gel using the Wizard SV Gel and PCR Clean-Up System Kit (Promega). The gel purified amplicons were cloned into pGEMT Easy Vector (Promega), and after sequence verification at the Australian Genome Research Facility (AGRF) PCR amplicons were diluted to 1 ng/μL and used as templates for a 1/10 standard curve serial dilution series. qRT-PCR was carried out in triplicate reactions in 96-well plates on the ViiA7 system (Applied Biosystems), each plate included RT negative and No template controls. Data were analyzed using Applied Biosystems Sequence Detection Software v1.2.3.

Nuclear-cytoplasmic miRNA target pulldown using ^{CNV}K oligos

For isolation of miR-29b associated transcripts in nucleus and cytoplasm, 9×10^6 NIH3T3 cells were transfected with 10 nM annealed oligonucleotides in 6-well format using Lipofectamine RNAiMax reagent (Life Technologies). Twenty-four hours post-transfection, cells were topped up with 2 mL complete growth media and allowed to recover for 8 h. Cells were rinsed with ice-cold PBS once, 500 μL PBS was added per well, and live cells were irradiated with 365 nm UV light for 10 min under a 20 W, 365 nm UV lamp at a distance of ~10 cm from the light source. Nuclear-cytoplasmic fractionation was carried out as described above, except using a slightly modified buffer A recipe (10 mM Tris-Cl pH 8, 140 mM NaCl, 1.5 mM MgCl$_2$, 0.5% NP-40 supplemented with 5 mM DTT and 1× Complete Mini Protease Inhibitor Cocktail [Roche]). To the nuclear pellet, 130 μL of sonication buffer (0.1% SDS, 1 mM EDTA, 10 mM Tris-Cl pH 8, 100 mM NaCl) was added and nuclei were lysed using a focused-ultrasonicator (Covaris S2, SonoLAB Single) using five pulses of 60 sec each with the following operating conditions: duty cycle: 2%; intensity: 3; cycles/burst: 200; bath temperature: 4°C; mode: frequency sweeping; degassing mode: continuous; volume: 130 μL in Covaris microTUBE, AFA fiber with Snap-Cap.

The volume of sonicated nuclear lysate and cytoplasmic supernatant was made up to 1 mL with modified buffer A such that the final NaCl concentration is 1 M. One-twentieth of the volume of lysate was kept as Input for each subcellular fraction. Sixty units/mL SUPERase•In (Ambion) was finally added to both nuclear and cytoplasmic lysates before isolating ^{CNV}K bound transcripts using streptavidin-coated Dynabeads. Dynabeads MyOne Streptavidin C1 (12 μL beads/million cells; Life Technologies) were preblocked with 1 μg/μL bovine serum albumin (Sigma) and 1 μg/μL yeast tRNA (Sigma), and incubated with salt-adjusted, nuclear-cytoplasmic lysates for 30 min at room temperature. Beads were then washed with hypotonic lysis buffer supplemented with 1 M NaCl before extracting associated RNAs from pulldown and Input samples using Trizol LS (Life Technologies) (supplemented with 17 μg/mL Glycoblue [Ambion] and 10 mM MgCl$_2$) as per manufacturer's instructions. Isolated RNA was resuspended in water and RNA quality and yield were assessed on the Agilent 2100 Bioanalyzer using the RNA 6000 Pico Chip Kit (Agilent).

Immunoblotting

For analysis of proteins associated with ^{CNV}K oligos and to analyze the purity of subcellular fractions by immunoblotting, the proteins were separated on a pre-cast 4%–20% gradient SDS-PAGE gel (Bio-Rad). The separated proteins were transferred onto methanol-activated polyvinylidene diflouride (PVDF) membranes and were blocked with 5% Skim milk (5% BSA was used for ab32381, Abcam) in 1× TBST buffer (50 mM Tris-Cl pH 7.6, 150 mM NaCl, 0.1% Tween-20). Membranes were then probed for specific antigens using antibodies against AGO2 (ab57113 for HeLa [Abcam]; ab32381 for NIH3T3 [Abcam], GAPDH [2275-PC-100, Trevigen], Lamin A/C [4777S, Cell Signalling], H3 Histone [9715S, Cell Signalling], or Calnexin [ab140818, Abcam]). Horseradish peroxidase (HRP) conjugated α-Rabbit or α-mouse antibody (Cell Signalling) was used as secondary antibody. Amersham ECL Western Blotting Detection Reagents (GE Healthcare) was used for immunodetection. See Supplemental Figure 11 for full-length digital images of immunoblots.

Luciferase assays

To select 88 human 3′ UTRs for Switchgear LightSwitch luciferase assay validation (Switchgear Genomics), 827 ^{CNV}K-identified human targets were binned into eight sets, from lowest to highest enrichment based on their PE value as follows: 4.15 to 5.41 (104 genes), 5.41 to 6.31 (103 genes), 6.31 to 7.38 (103 genes), 7.38 to 8.69 (104 genes), 8.69 to 10.7 (103 genes), 10.7 to 14.2 (103 genes), 14.2 to 19.5 (103 genes), 19.5 to 74.3 (104 genes). Each bin contained roughly the same number of genes and 11 genes were selected from each bin for luciferase assay validation. Thirty-five of the 88 genes were conserved between human and mouse. A subset of previously known targets was chosen as a positive control for the luciferase assays, while an additional seven pLightSwitch reporters were used as negative controls, determined by a lack of predicted miR-29b binding sites. This negative control set included the empty vector, four constructs from non-miR-29b targeted genes (*ACTB, GAPDH, LDHA, PPIA*), and two controls containing a random, nongenic, nonconserved sequence in place of the 3′ UTR. An optimized synthetic target consisting of sequence repeats that are fully

complementary to miR-29b was used as a positive control. 96-well plates were seeded with 10,000 HeLa cells per well 18–24 h before transfection to achieve 80% confluence at the time of transfection. Each transfection included 0.3 μL of DharmaFECT DUO transfection reagent, 100 ng of 3′ UTR reporter, and sufficient mimic or nontargeting control miRNA to yield a final concentration of 50 nM in a total volume of 100 μL/well. Each construct was transfected in triplicate separately with either the mimic (hsa-miR-29b) or the nontargeting control. Cells were incubated at 37°C for 24 h post-transfection before being harvested. One-hundred microliters of LightSwitch Luciferase Assay Reagent (Active Motif) were added to each well, plates were incubated at room temperature for 30 min and finally read on a SpectraMax L luminometer. 3′ UTR negative controls were used to normalize luminescence values, as per standard Switchgear normalization procedures.

Switchgear-provided UTR sequences were mapped to the genome with LASTZ and intersected with TargetScan predictions, revealing that 34/88 had no overlap with TargetScan predicted binding sites. Switchgear constructs include 3′ UTRs based on RefSeq annotations, while we utilized GENCODE annotations for RNA-seq mapping and enrichment analysis. Site type for a particular gene was defined as the highest confidence type of TargetScan predicted site in the UTR of that gene, i.e., if a UTR had a 7-mer-A1 and 8-mer site it was annotated as "8-mer." Target site efficacy was ordered as follows, i.e., 8mer > 7mer-m8 > 7mer-A1 > 6mer (Grimson et al. 2007). Pearson's χ^2 tests were carried out using the chisq.test function of the R stats library (R Core Team 2014). VisualiZation (Supplemental Fig. 5) was performed using the ggplot2 (Wickham 2009) library.

For luciferase assay data analysis, a normalization factor (NF) was calculated for the miR-29b mimic (mim) and control (NTC) experiments, according to the following formulas:

$$NF_{mim} = \langle \frac{\overline{ACTB_{mim}}}{\overline{ACTB_{NTC}} + \overline{ACTB_{mim}}}, \frac{\overline{GAPDH_{mim}}}{\overline{GAPDH_{NTC}} + \overline{GAPDH_{mim}}}, \frac{\overline{PPIA_{mim}}}{\overline{PPIA_{NTC}} + \overline{PPIA_{mim}}}, \\ \frac{\overline{LDHA_{mim}}}{\overline{LDHA_{NTC}} + \overline{LDHA_{mim}}}, \frac{\overline{Empty_{mim}}}{\overline{Empty_{NTC}} + \overline{Empty_{mim}}} \rangle$$

$$NF_{NTC} = \langle \frac{\overline{ACTB_{NTC}}}{\overline{ACTB_{NTC}} + \overline{ACTB_{mim}}}, \frac{\overline{GAPDH_{NTC}}}{\overline{GAPDH_{NTC}} + \overline{GAPDH_{mim}}}, \frac{\overline{PPIA_{NTC}}}{\overline{PPIA_{NTC}} + \overline{PPIA_{mim}}}, \\ \frac{\overline{LDHA_{NTC}}}{\overline{LDHA_{NTC}} + \overline{LDHA_{mim}}}, \frac{\overline{Empty_{NTC}}}{\overline{Empty_{NTC}} + \overline{Empty_{mim}}} \rangle$$

where gene$_{mim}$ and gene$_{NTC}$ are the average luminosity recorded in the three replicate miR-29b mimic and nontargeting control experiments, respectively. The luminosity of each interrogated 3′ UTR reporter was then divided by the corresponding normalization factor, and a Welch's t-test with a Benjamini–Hochberg (Benjamini and Hochberg 1995) false discovery rate of 0.05 was used to identify significant differences between miR-29b mimic and nontargeting control normalized luminosity.

Northern blot hybridization

To assess if cells equally take up both ^{CNV}K-miR-29b and ^{CNV}K-scram oligonucleotides, 6 million NIH3T3 cells were transfected with ^{CNV}K oligos. Crosslinking, pulldown, and RNA isolation were performed as described above. Ten micrograms of Input RNA was loaded per blot and the entire yield of pulldown RNA was used. tRNA-Lys was used to serve as an internal control.

To assess if transfected ^{CNV}K oligos are imported into the nucleus, 9 million NIH3T3 cells were transfected with 10 nM ^{CNV}K oligos

and nuclear, cytoplasmic pulldown RNA isolated as described above without 365 nm UV crosslinking. The entire yield of pulldown RNA was used for northern blot hybridization. Four micrograms of total RNA isolated from respective transfected cells was used as positive control. To demonstrate nuclear localization of miR-29b, cell equivalent amounts of NIH3T3 nuclear and cytoplasmic RNA was used for every blot. Ten micrograms of total RNA was loaded per blot to serve as an internal control.

An equal volume of 2× formamide loading dye was mixed with RNA and samples denatured at 95°C for 5 min. Fifteen percent polyacrylamide/8M urea/0.25X TBE gel was used to detect miRNAs. RNA was transferred onto a nylon membrane (Hybond-NX, Amersham Biosciences) in a semi-dry fashion using 0.25× TBE. The RNA was chemically cross-linked for 2 h at 55°C using 1-Ethyl-3-[3-dimethylaminopropyl]carbodiimide hydrochloride (EDC). The blots were hybridized with radiolabeled probes and the membrane exposed to storage phosphor screen for at least 3 h. Signal was detected using Typhoon 8600 variable mode imager (Amersham Biosciences).

Splinted ligation was performed as previously described (Maroney et al. 2007). Briefly, 2 μg RNA was mixed with 1 pmol of radiolabeled ligation oligo in the presence of 1 pmol of bridge oligonucleotide in a 10 μL reaction. The reaction was incubated at 94°C for 1 min, 65°C for 2 min, and 37°C for 10 min. The RNA and ligation oligo were ligated using the Ligate–IT Rapid ligation Kit (USB) and the mixture incubated at 30°C for 1 h. Finally, 1 μL of Shrimp Alkaline Phosphatase (USB) was added per reaction and incubated at 37°C for 15 min. All incubations were carried out in a thermacycler PCR machine. Samples were denatured in an equal volume of 2× formamide loading dye at 95°C for 3 min and loaded onto 15% PAGE 8 M urea/1× GTG buffer. The bridge oligos were synthesized with three-carbon spacers (SpC3) on either end, while the ligation oligo had one spacer at its 3′ end (Supplemental Table 6). The carbon spacers serve as unligatable groups, thus preventing the formation of oligo concatemers. All oligos were synthesized by IDT.

Library preparation and sequencing

RNA samples were DNase treated using TURBO DNase (Ambion) and purified using the RNeasy MinElute Cleanup Kit (Qiagen) according to manufacturer's instructions. The RNA was quantified using the Agilent RNA 6000 Pico Kit (Agilent Technologies), and a mix of oligo(dT) and random hexamers primers were used to prepare sequencing libraries from 5 ng RNA using the TotalScript RNA-seq Kit (Epicentre), as per manufacturer's instructions with a few modifications. Modifications in the library preparation protocol include the use of AxyPrep magnetic beads (AxyPrep Mag PCR Clean-up Kit 50 mL, Axygen) instead of AMPure XP magnetic beads throughput the protocol; 2× Phusion High Fidelity PCR Master Mix was replaced with a cocktail of 5× Phusion High Fidelity PCR Master Mix HF Buffer (NEB), 200 μM dNTPs and 0.5 units Phusion High-Fidelity DNA Polymerase (NEB); finally, a 1:1 ratio of magnetic beads to gap-filled amplicon was used in the final clean-up step of the protocol. Libraries were sequenced on an Illumina HiSeq 2500 platform (QCMG) to obtain 100 bp paired-end reads.

For small RNA deep sequencing analysis, nuclear and cytoplasmic RNA was isolated from NIH3T3 cells using the fractionation protocol

described above. These samples were used to construct two small RNA libraries using the Illumina TruSeq Small RNA Kit (Illumina) as per the manufacturer's instructions. The libraries were sequenced on an Illumina HiSeq 2000 platform (QCMG) to obtain 50 bp single-end reads. The small RNA reads were trimmed, adaptors removed, and identical reads collapsed to a single read using the FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/index.html) commands "fastx_clipper -a TGGAATTCTC -l 16 -i <input.fastq> -o <clipped.fastq> -Q33" and "fastx_collapser -i <clipped.fastq> -o <clipped.collapsed.fasta> -Q33". The final reads were mapped against the UCSC mouse reference genome (mm10) using bowtie 2 (Langmead and Salzberg 2012) with settings: "--local -k 100 -f -L 10 -N 1 -i L,1,0". The resulting SAM file was converted to BAM format using SAMtools (Li et al. 2009a), and sorted by both coordinate and read name using picard (http://broadinstitute.github.io/picard). Reads in name-sorted BAM overlapping annotated miRNA genes were counted using a custom Python script. All scripts are available from the authors by request.

Read mapping and quantification

STAR (Dobin et al. 2013) version 2.3.0 was used to map reads to the genome (mm10 for mouse and hg19 for human). GENCODE annotations (version M2 for mouse and 19 for human) were used to construct the splice junction database and as a reference for the count tables. Gene-level count tables were obtained using the HTSeq framework in union mode (Anders et al. 2015). Count tables were filtered to discard lowly expressed genes, defined as those with fewer than a total of 30 tags mapping across all of the input data sets. A pseudocount of one was added to each observation of each gene, and counts were then normalized to the number of uniquely mapped reads in each of the sequencing libraries. A pulldown enrichment (PE) value was determined for every set of four experiments:

$$PE = (29b_{pulldown} - avg(29b_{input}, scrambled_{input})) / scrambled_{pulldown}$$

Genes were considered enriched if the PE of all replicates was greater than four.

Quality control and exploratory data analysis

DESeq2 (Love et al. 2014) was used for exploratory data analysis. Hierarchical clustering, correlation, and overlap analysis of mouse and human data sets (Supplemental Fig. 4) revealed good concordance between similar samples in the repetitions of the experimental series, and between genes considered to be miR-29b targets as a result of having a PE > 4. Few differences were observed between samples prepared using the intact and lysed protocols.

Gene ontology analysis

Gene ontology, pathway, transcription factor binding site, and miRNA target site enrichment analysis were carried out using the ToppGene server (Chen et al. 2009) with a 0.05 Benjamini–Hochberg false discovery rate and requiring a minimum of two features counted in the test set. For mouse genes, the MGI name was used as input; this resulted in 522 unique gene names queried. For human genes, 763 genes had HGNC gene names and were

present in the ToppGene database. In order to incorporate ontology annotations into the list of genes tested using SwitchGear luciferase assays in Supplemental Figure 5, the gene list was submitted to ToppGene, and genes that were part of the "Cell cycle," "DNA Damage Response," "Chromatin," "Extracellular matrix and collagens" (which represents the union of "Extracellular matrix" and "Collagens" terms), and "Occluding junctions" ontologies were annotated as such (Supplemental Table 3).

Comparison of ^{CNV}K-identified genes with previously reported miR-29b targets

The list of ^{CNV}K-identified genes was compared to a list of previously reported miR-29b targets (Supplemental Table 1), taking into consideration levels of expression of these genes in the cell lines used. This enabled us to take into account the fact that the ^{CNV}K approach cannot identify a gene to be a target of miR-29b if that gene is not expressed. Of the 69 known mouse targets (Supplemental Table 4), 55 were tested for differential enrichment since they passed the greater than 30 counts across all NIH3T3 input data sets thresholds. Similarly, 83/111 known human targets (Supplemental Table 4) were interrogated.

Comparison with predicted miR-29b targets

TargetScan predictions were downloaded from the UCSC Genome Browser Table Browser ("TargetScan miRNA Regulatory Sites" track; Supplemental Table 7). Predictions were subdivided into four categories: those containing a conserved 8-mer site ("8mer"), those with a context score greater than 80 ("GT8ctx"), those that have both a conserved 8-mer site and a context score greater that 80 ("high-confidence"), and all sites ("total").

Mouse targets

Predictions were intersected with ensembl transcripts using the UCSC Genome Browser Table Browser functionality (Rosenbloom et al. 2015). BiomaRt (Durinck et al. 2009) was used to retrieve ENSMG gene identifiers for the ENSMT transcripts, and as most of these were represented in GENCODEM2, the corresponding ENSMUSG was used. Identifiers for those not present were retrieved using biomaRt of the ensembl archive version May 2012, and all but three ENSMUSG gene names were present in the current assembly. Those three genes were manually queried and current identifiers recovered: ENSMUSG00000079828 (not in primary assembly, BLAST of the peptide sequence against the "nr" database showed 100% identity with ENSMUSG00000093930), ENSMUSG00000091454 (merged into ENSMUSG00000095677 record as of Ensembl 68), and ENSMUSG00000091645 (only present in release 67, no peptide product; visualizing this gene in UCSC shows overlap with Zfp36l1, which now has the identifier). A total of 743 TargetScan miR-29b sites were annotated in UCSC, and these overlapped with 1486 ensembl transcripts, corresponding to 664 ensembl mouse genes (ENSMG) (13 sites had no overlap with an ensembl transcript); of these 664, 509 were expressed above the 30+ tags threshold across all four input data sets and hence considered for potential enrichment. A total of 226 8-mer miR-29b sites were identified, which overlapped with 449 ensembl transcripts, corresponding to 203

ensembl mouse genes (ENSMG) (six sites had no overlap with an ensembl transcript); of these 203, 148 were expressed above the 30 + tags threshold across all four input data sets and hence considered for potential enrichment. A total of 263 TargetScan miR-29b sites had a context score ≥80, which overlapped with 516 ensembl transcripts, corresponding to 234 ensembl mouse genes (ENSMG) (four sites had no overlap with an ensembl transcript); of these 234, 167 were expressed above the 30+ tags threshold across all four input data sets and hence considered for potential enrichment. A total of 170 TargetScan miR-29b sites had a context score ≥80 and an 8-mer sequence, which corresponded to 153 ensembl mouse genes (ENSMG); of these, 106 were expressed above the 30+ tags threshold across all four input data sets and hence considered for potential enrichment.

Human targets

Predictions were intersected with GENCODE 19 Comprehensive transcripts using the UCSC Genome Browser Table Browser functionality. BiomaRt was used to retrieve ENSG gene identifiers for the ENST transcripts. A total of 948 miR-29b sites were identified, which overlapped with 2235 transcripts, corresponding to 888 genes (ENSG) (four sites had no overlap with an annotated transcript); of these 888, 669 were expressed above the 30+ tags threshold across all three input data sets and hence considered for potential enrichment. A total of 272 8-mer miR-29b sites were identified, which overlapped with 623 transcripts, corresponding to 268 human genes (ENSG) (one site had no overlap with an annotated transcript); of these 268, 202 were expressed above the 30+ tags threshold across all three input data sets and hence considered for potential enrichment. A total of 311 TargetScan miR-29b sites had a context score ≥80, which overlapped with 657 transcripts, corresponding to 299 human genes (ENSG) (one site had no overlap with an annotated transcript); of these 299, 218 were expressed above the 30+ tags threshold across all three input data sets and hence considered for potential enrichment. A total of 186 TargetScan miR-29b sites had a context score ≥80 and an 8-mer sequence, which overlapped with 421 transcripts, corresponding to 191 human genes (ENSG); of these 191, 136 were expressed above the 30+ tags threshold across all three input data sets and hence considered for potential enrichment.

Permutation testing

A total of 567 (mouse) and 827 (human) miR-29b target hits were identified based on the PE metric filtering described above. We used the following random permutation model to assess whether the high degree of overlap with TargetScan predictions occurred by chance alone. We generated 10,000 random sets of 567 and 827 genes, selecting from those expressed (>30 reads) in the input mouse and human data sets, respectively. We then counted how many of these genes were represented in the TargetScan predictions, with a context score >80, predicted to have an 8-mer binding site or both of these. A *P*-value for each comparison was calculated by using the R statmod (Smyth et al. 2013) library using the approach described by Phipson (Phipson and Smyth 2010), which considers the number of times the number of overlaps between a random set and the TargetScan predictions exceeded that of our ^{CNV}K-detected target set. A total of one (mouse) and 10 (human) genes were characterized

as enriched using a PE metric for the scrambled oligonucleotide, none of which were predicted to be miR-29b targets by TargetScan.

Target conservation analysis

BiomaRt was used to identify human orthologs of mouse genes based on the cross-species mappings for GENCODE 19 and GENCODE M2. A total of 542/567 genes identified as miR-29b targets in NIH3T3 cells had annotated human orthologs, of which 483 were expressed at levels of greater than 30 tags in HeLa cells across all of the input data sets and hence tested for pulldown enrichment. Of the 827 identified human hits, 656 were expressed above the minimum tag threshold in mouse (total of at least 30 counts across all eight input data sets) and hence tested for enrichment. The VennDiagram and ggplot2 R package was used to visualize the results.

SUPPLEMENTAL MATERIAL

Supplemental material is available for this article.

REFERENCES

Anders S, Pyl PT, Huber W. 2015. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31:** 166–169.

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B Methodol* **57:** 289–300.

Chen J, Bardes EE, Aronow BJ, Jegga AG. 2009. ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic Acids Res* **37:** W305–W311.

Chou J, Lin JH, Brenot A, Kim JW, Provot S, Werb Z. 2013. GATA3 suppresses metastasis and modulates the tumour microenvironment by regulating microRNA-29b expression. *Nat Cell Biol* **15:** 201–213.

Cloonan N, Wani S, Xu Q, Gu J, Lea K, Heater S, Barbacioru C, Steptoe AL, Martin HC, Nourbakhsh E. 2011. MicroRNAs and their isomiRs function cooperatively to target common biological pathways. *Genome Biol* **12:** R126.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29:** 15–21.

Durinck S, Spellman PT, Birney E, Huber W. 2009. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat Protoc* **4:** 1184.

Fujimoto K, Hiratsuka-Konishi K, Sakamoto T, Ohtake T, Shinohara K, Yoshimura Y. 2012. Specific and reversible photochemical labeling of plasmid DNA using photoresponsive oligonucleotides containing 3-cyanovinylcarbazole. *Mol Biosyst* **8:** 491–494.

Gagnon KT, Li L, Chu Y, Janowski BA, Corey DR. 2014. RNAi factors are present and active in human cell nuclei. *Cell Rep* **6:** 211–221.

Grimson A, Farh KKH, Johnston WK, Garrett-Engele P, Lim LP, Bartel DP. 2007. MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol Cell* **27:** 91–105.

Hafner M, Landthaler M, Burger L, Khorshid M, Hausser J, Berninger P, Rothballer A, Ascano M Jr, Jungkamp AC, Munschauer M, et al. 2010. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* **141:** 129–141.

Helwak A, Kudla G, Dudnakova T, Tollervey D. 2013. Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding. *Cell* **153:** 654–665.

Hwang HW, Wentzel EA, Mendell JT. 2007. A hexanucleotide element directs microRNA nuclear import. *Science* **315:** 97–100.

Jeffries CD, Fried HM, Perkins DO. 2010. Additional layers of gene regulatory complexity from recently discovered microRNA mechanisms. *Int J Biochem Cell Biol* **42:** 1236–1242.

Konig J, Zarnack K, Rot G, Curk T, Kayikci M, Zupan B, Turner DJ, Luscombe NM, Ule J. 2010. iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. *Nat Struct Mol Biol* **17:** 909–915.

Kriegel AJ, Liu Y, Fang Y, Ding X, Liang M. 2012. The miR-29 family: genomics, cell biology, and relevance to renal and cardiovascular injury. *Physiol Genomics* **44:** 237–244.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9:** 357–359.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009a. The sequence alignment/map format and SAMtools. *Bioinformatics* **25:** 2078–2079.

Li Z, Hassan MQ, Jafferji M, Aqeilan RI, Garzon R, Croce CM, Van Wijnen AJ, Stein JL, Stein GS, Lian JB. 2009b. Biological functions of miR-29b contribute to positive regulation of osteoblast differentiation. *J Biol Chem* **284:** 15676–15684.

Li Y, Wang H, Tao K, Xiao Q, Huang Z, Zhong L, Cao W, Wen J, Feng W. 2013. miR-29b suppresses CML cell proliferation and induces apoptosis via regulation of BCR/ABL1 protein. *Exp Cell Res* **319:** 1094–1101.

Liao JY, Ma LM, Guo YH, Zhang YC, Zhou H, Shao P, Chen YQ, Qu LH. 2010. Deep sequencing of human nuclear and cytoplasmic small RNAs reveals an unexpectedly complex subcellular distribution of miRNAs and tRNA 3′ trailers. *PLoS One* **5:** e10563.

Licatalosi DD, Mele A, Fak JJ, Ule J, Kayikci M, Chi SW, Clark TA, Schweitzer AC, Blume JE, Wang X, et al. 2008. HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature* **456:** 464–469.

Loeb GB, Khan AA, Canner D, Hiatt JB, Shendure J, Darnell RB, Leslie CS, Rudensky AY. 2012. Transcriptome-wide miR-155 binding map reveals widespread noncanonical microRNA targeting. *Mol Cell* **48:** 760–770.

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15:** 550.

Magnusson M, Brun AC, Miyake N, Larsson J, Ehinger M, Bjornsson JM, Wutz A, Sigvardsson M, Karlsson S. 2007. HOXA10 is a critical regulator for hematopoietic stem cells and erythroid/megakaryocyte development. *Blood* **109:** 3687–3696.

Maroney PA, Chamnongpol S, Souret F, Nilsen TW. 2007. A rapid, quantitative assay for direct detection of microRNAs and other small RNAs using splinted ligation. *RNA* **13:** 930–936.

Mcdonough WS, Tran NL, Berens ME. 2005. Regulation of glioma cell migration by serine-phosphorylated P311. *Neoplasia* **7:** 862–872.

Park SY, Lee JH, Ha M, Nam JW, Kim VN. 2009. miR-29 miRNAs activate p53 by targeting p85 α and CDC42. *Nat Struct Mol Biol* **16:** 23–29.

Peng X, Yuan S, Tan J, Ma B, Bian X, Xu C, He W, Cao H, Huang Z, Cui Y, et al. 2012. Identification of ITGB4BP as a new interaction protein of P311. *Life Sci* **90:** 585–590.

Phipson B, Smyth GK. 2010. Permutation *P*-values should never be zero: calculating exact *P*-values when permutations are randomly drawn. *Stat Appl Genet Mol Biol* **9:** Article39.

R Core Team. 2014. *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.

Rosenbloom KR, Armstrong J, Barber GP, Casper J, Clawson H, Diekhans M, Dreszer TR, Fujita PA, Guruvadoo L, Haeussler M. 2015. The UCSC Genome Browser database: 2015 update. *Nucleic Acids Res* **43:** D670–D681.

Sakamoto T, Shigeno A, Ohtaki Y, Fujimoto K. 2014. Photo-regulation of constitutive gene expression in living cells by using ultrafast photo-cross-linking oligonucleotides. *Biomater Sci* **2:** 1154–1157.

Smyth G, Hu Y, Dunn P, Phipson B, Chen Y. 2013. *statmod: statistical modeling*. R package version **1:** 17.

Suh EJ, Remillard MY, Legesse-Miller A, Johnson EL, Lemons JM, Chapman TR, Forman JJ, Kojima M, Silberman ES, Coller HA. 2012. A microRNA network regulates proliferative timing and extracellular matrix synthesis during cellular quiescence in fibroblasts. *Genome Biol* **13:** R121.

Wang B, Li W, Liu H, Yang L, Liao Q, Cui S, Wang H, Zhao L. 2014. miR-29b suppresses tumor growth and metastasis in colorectal cancer via downregulating Tiam1 expression and inhibiting epithelial–mesenchymal transition. *Cell Death Dis* **5:** e1335.

Wickham H. 2009. *ggplot2: elegant graphics for data analysis.* Springer, New York.

Yoshimura Y, Fujimoto K. 2008. Ultrafast reversible photo-cross-linking reaction: toward in situ DNA manipulation. *Org Lett* **10:** 3227–3230.

Yue MM, Lv K, Meredith SC, Martindale JL, Gorospe M, Schuger L. 2014. Novel RNA-binding protein P311 binds eukaryotic translation initiation factor 3 subunit b (eIF3b) to promote translation of transforming growth factor β1-3 (TGF-β1–3). *J Biol Chem* **289:** 33971–33983.

Zhang Y, Huang XR, Wei LH, Chung AC, Yu CM, Lan HY. 2014. miR-29b as a therapeutic agent for angiotensin II-induced cardiac fibrosis by targeting TGF-β/Smad3 signaling. *Mol Ther* **22:** 974–985.