RISC RNA Sequencing for Context-Specific Identification of In Vivo MicroRNA Targets

Scot J. Matkovich, Derek J. Van Booven, William H. Eschenbacher, Gerald W. Dorn II

Rationale: MicroRNAs (miRs) are expanding our understanding of cardiac disease and have the potential to transform cardiovascular therapeutics. One miR can target hundreds of individual mRNAs, but existing methodologies are not sufficient to accurately and comprehensively identify these mRNA targets in vivo.

Objective: To develop methods permitting identification of in vivo miR targets in an unbiased manner, using massively parallel sequencing of mouse cardiac transcriptomes in combination with sequencing of mRNA associated with mouse cardiac RNA-induced silencing complexes (RISCs).

Methods and Results: We optimized techniques for expression profiling small amounts of RNA without introducing amplification bias and applied this to anti–Argonaute 2 immunoprecipitated RISCs (RISC-Seq) from mouse hearts. By comparing RNA-sequencing results of cardiac RISC and transcriptome from the same individual hearts, we defined 1645 mRNAs consistently targeted to mouse cardiac RISCs. We used this approach in hearts overexpressing miRs from Myh6 promoter-driven precursors (programmed RISC-Seq) to identify 209 in vivo targets of miR-133a and 81 in vivo targets of miR-499. Consistent with the fact that miR-133a and miR-499 have widely differing “seed” sequences and belong to different miR families, only 6 targets were common to miR-133a– and miR-499–programmed hearts.

Conclusions: RISC-sequencing is a highly sensitive method for general RISC profiling and individual miR target identification in biological context and is applicable to any tissue and any disease state. (Circ Res. 2011;108:00-00.)

Key Words: RNA-induced silencing complex ■ RNA sequencing ■ mRNA ■ miRNA ■ miR-133a ■ miR-499

MicroRNAs (miRs) are small, noncoding RNAs that regulate critical aspects of cell function including growth, differentiation, and programmed death. These ~22-nucleotide single-stranded oligomers bind through Watson–Crick base-pairing to complementary sequences in mRNA targets, thereby accelerating mRNA degradation or interfering with protein translation. miRs play essential roles in tissue development and homeostasis, and their capacity to act as mediators of disease have been demonstrated in multiple conditions and organ systems.1–3

The effects of miRs accrue from their binding to specific mRNA targets. Bioinformatics analysis has suggested that 30% of all mRNAs may be regulated by miRs. Because a single miR can bind to dozens or hundreds of different mRNAs, it has the potential to orchestrate complex gene networks and equally complicated phenotypes. For this reason, there is tremendous interest in developing miR antagonists and/or mimetics for clinical therapeutics.4 Before the therapeutic potential of miRs can be fulfilled, however, the full spectrum of their specific mRNA targets needs to be determined in the context of the organ system and pathophysiological processes of interest. Achieving this goal has lagged behind experimental miR manipulation and characterization of resulting phenotypes.5 One approach for identifying miR targets compares miR and mRNA expression signatures to detect reciprocally regulated miR-mRNA pairs.6 This approach is limited because only a fraction of miR effects are mediated through mRNA destabilization (the remainder by translational suppression)7 and because indirect miR effects can regulate mRNAs that are not direct targets.8 Bioinformatics have also been used to identify putative mRNA targets based on miR antisense sequence pairing to mRNA 3′ untranslated regions (UTRs), evolutionary conservation, and the predicted binding energy of miR-mRNA duplexes. However, different bioinformatics platforms differ markedly in their predictions,9 and even if the predictions are accurate, bioinformatics does not provide information about biological context. Thus, techniques are needed for comprehensive, unbiased in vivo identification of miR targets in health and disease.

An approach of isolating mRNAs sequestered by miRs into RNA-induced silencing complexes (RISCs) has been used to...
profile miR-targeted mRNAs in cultured cells and mouse brains.\textsuperscript{10,11} Although conceptually attractive, this technique has been limited in practice by use of microarrays requiring RNA amplification, which is a potential source of experimental error. Here, we reveal the extent to which RNA amplification biases transcriptional profiling. To avoid this confounder, we describe a technique for annotation and quantitation of mRNAs in the RNA-induced sequencing complexes (the “RISCome”) of mouse hearts using next generation massively parallel sequencing of unamplified mRNAs isolated from myocardial Argonaute (Ago)2 immune complexes. By “programming” the cardiomyocytes with excess miR, we then applied the technique to identification of cardiomyocyte-autonomous mRNA targets for 2 structurally and functionally distinct cardiac-expressed miRs, miR-133a, and miR-499. These procedures are applicable to any miR in any tissue and will complement mRNA footprinting\textsuperscript{11} and proteomics analyses\textsuperscript{12} to predict in vivo effects of proposed miR-based therapeutics.

### Methods

#### Preparation and Quantification of Total Myocardial RNA

Cardiac miR-133a transgenic mice have been described previously.\textsuperscript{13} miR-499 transgenic mice were generated similarly by cloning a 530-bp fragment surrounding the mature mouse miR-499 sequence (chromosome 2; intron 19 of Myh6) into the Myh6 promoter construct. Total cardiac RNA was isolated from flash-frozen ventricular apices using the Dynabead mRNA purification system (Invitrogen). A total of 200 ng of cardiac mRNA was fragmented to \( \sim 200 \) nucleotides by heating to 94°C for 2.5 minutes in 40 mM Tris acetate pH 8.2, 100 mM sodium acetate, 30 mM magnesium acetate, and immediately chilled on ice. After purification on Ambion NucAway columns, 100 ng of fragmented cardiac mRNA was reverse-transcribed using SuperScript III (Invitrogen) with random hexamers according to the directions of the manufacturer (30 minutes, 50°C), followed by second-strand cDNA synthesis for 2 hours, 16°C, in 20 mM Tris-HCl pH 6.9, 90 mM KCl, 4.6 mM MgCl\(_2\), 150 mM sodium chloride, 200 U/\( \mu \)L M-MLV reverse transcriptase, 200 U/\( \mu \)L E. coli DNA ligase, 0.27 U/\( \mu \)L E. coli DNA polymerase I, and 0.013 U/\( \mu \)L E. coli RNase H (New England Biolabs).

#### Reverse Transcription and Preparation of cDNA

Preparation of cDNA fragments was modified from previously described protocols.\textsuperscript{8,14} For poly(A)\(^+\) RNA, 4 \( \mu \)g of total cardiac RNA isolated from ventricular apices was twice oligo(dT)-selected using the Dynabead mRNA purification system (Invitrogen). A total of 200 ng of cardiac mRNA was fragmented to \( \sim 200 \) nucleotides by heating to 94°C for 2.5 minutes in 40 mM Tris acetate pH 8.2, 100 mM sodium acetate, 30 mM magnesium acetate, and immediately chilled on ice. After purification on Ambion NucAway columns, 100 ng of fragmented cardiac mRNA was reverse-transcribed using SuperScript III (Invitrogen) with random hexamers according to the directions of the manufacturer (30 minutes, 50°C), followed by second-strand cDNA synthesis for 2 hours, 16°C, in 20 mM Tris-HCl pH 6.9, 90 mM KCl, 4.6 mM MgCl\(_2\), 150 mM sodium chloride, 200 U/\( \mu \)L M-MLV reverse transcriptase, 200 U/\( \mu \)L E. coli DNA ligase, 0.27 U/\( \mu \)L E. coli DNA polymerase I, and 0.013 U/\( \mu \)L E. coli RNase H (New England Biolabs).

#### Integrated Reverse Transcription and Amplification of Low-Input RNA

RNA amplification before fragmentation and sequencing library construction was performed according to Mitreva and Mardis.\textsuperscript{15} Briefly, polyA-tailed RNA was selected from total RNA using a primer recognizing 5’ mRNA ends and a primer containing an oligo(dT)\(_{15}\) moiety. Following reverse transcription, a third primer recognizing common elements in the first 2 primers was used to amplify cDNAs and to add biotin tags (primer sequences as described in\textsuperscript{15}). Optimization of amplification cycles was performed as described,\textsuperscript{15} and biotin tags were removed on streptavidin beads following restriction enzyme digestion.

### Immunoprecipitation of Mouse Cardiac Ago2-Associated RNA and Preparation of cDNA

Conditions for Ago2 immunoprecipitation were adapted from Karginov et al.\textsuperscript{10} Frozen mouse heart bases were homogenized in 500 \( \mu \)L of ice-cold 50 mM Tris-HCl, 5 mM EDTA, 5 mM EGTA, pH 7.5, with Roche Complete protease inhibitors. Yeast tRNA (Invitrogen) and SUPERase-IN (Ambion) were added to final concentrations of 1 mg/mL and 1 U/\( \mu \)L, respectively, and unbroken cellular material was removed at 100g, 5 minutes, 4°C. Nonidet P-40 was added to a final concentration of 0.5% (wt/vol) to solubilize proteins (15 minutes, 4°C; rotating) and insoluble material was removed at 10 000g, 15 minutes, 4°C. The supernatant was added to 50 \( \mu \)L of protein G–coupled Dynabeads (Invitrogen), to which 5 \( \mu \)g of antimouse Ago2 monoclonal antibody (Wako Pure, clone no. 2D4, lot PEM0820) had been previously bound according to the Dynabead protocol. Following 1 hour of rotational incubation at 4°C, the beads were washed 3 times with Dynabead washing buffer, transferring the suspension to a fresh tube for the last wash. Beads were pelleted from the remaining suspension, the supernatant was removed, and 500 \( \mu \)L of TRizol was added to the immunoprecipitated material to extract RNA. Ago2 immunoprecipitate–associated RNA was fragmented in acetate buffer without poly-A selection, purified on NucAway columns, and one-half was used for cDNA synthesis using the protocol described above.

#### Construction of DNA-Encoded Short-Read Libraries for Illumina Sequencing

Detailed methods for preparation of Illumina sequencing libraries from mouse cardiac RNA have been described recently.\textsuperscript{8} Briefly, cDNAs were end-repaired and 3’ A-overhangs added. Illumina adapters with T-overhangs and customized to include 3-nucleotide “barcodes” were ligated to the cDNA at 10:1 molar excess and DNA in the 200- to 400-bp range was isolated via gel purification (Qiagen) on 2% low-melting agarose. One-third of the gel-purified material was amplified with 12 cycles (total myocardial cDNA) or 16 cycles (RISC-immunoprecipitated cDNA) of Phusion polymerase–mediated (New England Biolabs, no. F531) PCR (10 seconds, 98°C; 30 seconds, 65°C; 30 seconds, 72°C; followed by final 5 minutes at 72°C), using oligonucleotides complementary to Illumina sequencing adapters. The final, amplified libraries were again column-purified and quantified using PicoGreen (Quant-It, Invitrogen).

Four barcoded libraries were combined in equimolar (10 nmol/L) amounts and diluted to 6 pmol/L for cluster formation on a single Illumina Genome Analyzer II flow cell lane, followed by single-end sequencing. Base calling, library sorting by barcode, and mapping to the transcriptome were performed as described previously,\textsuperscript{8} using updated versions of the software packages Bowtie (release 0.12.3),\textsuperscript{16} TopHat (release 1.0.13),\textsuperscript{17} and Cufflinks (release 0.8.1).\textsuperscript{18} Cufflinks outputs gene expression values in terms of FPKM (fragments per kilobase of exon per million mapped reads).\textsuperscript{18} When performing outputs gene expression values in terms of FPKM (fragments per kilobase of exon per million mapped reads), in which an RPKM of 3 corresponds to 1 copy per cell in cardiac samples.\textsuperscript{8,14} We used the default options supplied with these software packages in our analyses and analyzed only those RNA elements that had expression signals in at least 2 of 4 biological replicates.
Computation of RISC Abundance and Enrichment Scores

Because they are not poly-A–selected, RISCome sequencing libraries contained a higher proportion of ribosomal RNA (32% to 42%) than poly-A–selected transcriptome libraries (3% to 5%). To permit comparison between the transcriptome and RISCome, measures of RNA abundance expressed as FPKM were adjusted for the proportion of mRNA. RISC enrichment scores to assess RISC enrichment or depletion were defined as (RISCome-adjusted FPKM/transcriptome-adjusted FPKM) of ≥2.0 or ≤0.5; P < 0.0001; FDR < 0.01. Scores of <1 were plotted as −1 per score for clarity.

Luciferase Reporter Constructs

Full (if ≥1 kb) or partial (regions selected to contain miR-133a binding “seeds”) 3' UTRs of candidate miR-133a target genes were cloned into the dual-luciferase reporter vector psiCheck2 (Promega). The miR-133a genomic precursor was the same as that used to create miR-133a transgenic mice. A total of 100 ng of psiCheck2 construct, together with 100 ng of pcDNA3.1 or pcDNA3.1+miR-133a genomic precursor, was transfected in triplicate into HEK293 cells on 24-well plates using Fugene HD (Promega). After 48 hours, duplicate determinations of Renilla and firefly luciferase activities were performed after harvesting cells into 120 L of Glo Lysis Buffer (Promega). Each lysate (30 μL) was analyzed using Dual-Glo Luciferase reagents (Promega) on a SpectraMax M5e plate reader (Molecular Devices). Renilla 3' UTR–coupled luciferase activity was normalized to constitutive firefly luciferase activity for each well.

Statistical Analysis

Gene symbols and FPKM values were imported into Partek Genomics Suite (version 6.5; Partek, St Louis, Mo) for comparison of RISCome and transcriptome expression values, computation of probability values and false discovery rates, and preparation of probability value versus fold change (volcano) plots. GraphPad Prism was used for χ2 tests. Luciferase activities were compared using Student t test (GraphPad Prism). Gene ontology overrepresentation was performed using a hypergeometric test with Benjamini–Hochberg false discovery rate correction. Significant differences were defined as P < 0.05 unless otherwise described.

Results

RNA Amplification Biases mRNA Expression Signatures

miR-mRNA pairings in any tissue depend on at least 3 factors: the identity and quantity of expressed miRs, the identity and quantity of expressed mRNAs, and the potential for those miRs and mRNAs to interact via sequence complementarity. The former 2 factors vary with tissue type and pathological circumstance. It is therefore necessary to profile miR-mRNA interactions in the proper biological context.

Previous RNA profiling of RISC-associated mRNA used expression microarrays, which require RNA amplification that might alter the apparent transcriptional signature. To determine the effect of mRNA amplification on transcript profiling, we compared mRNA signatures of amplified and nonamplified mRNA from the same mouse heart samples. A total of 50 ng of unamplified total mouse cardiac RNA underwent 2 poly-A– selections to yield 150 ng of mRNA for reverse transcription, sequencing library production, and RNA sequencing on an Illumina Genome Analyzer II. In parallel, 50 ng of the same total mouse cardiac RNA (ie, 1/100th the above initial input amount) was poly-A–selected, reverse-transcribed, and amplified using an integrated multistage procedure developed specifically for expression profiling of limited RNA samples, and Illumina sequencing libraries were prepared (Figure 1, left pathway). Sequencing
The RISC consists of Ago1 and -2 proteins, TNRC6 family members, miRs, and their mRNA targets. After identifying an antibody that immunoprecipitates endogenous Ago2 from mouse hearts (Figure 2a), we determined whether endogenous cardiac Ago2 immunoprecipitates contained sufficient RISC-associated cardiac mRNA for RNA sequence analysis without sample crosslinking, amplification, or pooling previously used with Ago2 pull-down. Ventricular tissue from four 8-week-old male FVB/N mouse hearts was transversely sectioned; one-third of each heart was used for mRNA sequencing to characterize the overall cardiac transcriptomes, and the remaining two-thirds were used for Ago2 immunoprecipitation, RNA extraction, and RISCome RNA-sequencing analysis. Although there was 96.5% overlap in the individual mRNAs represented in the cardiac transcriptome and RISCome (9779 of 894 transcriptome mRNAs and 10 690 ± 205 RISCome mRNAs; P = 0.31), 1645 individual mRNAs (15% of the cardiac transcriptome) were significantly RISC-enriched in the Ago2 immunoprecipitates (Figure 2b). RISC enrichment or depletion was defined as (RISCome expression/transcriptome expression) of ≥2.0 or ≤0.5; P < 0.0001; FDR < 0.01. The enriched transcripts define
the subset of cardiac-expressed mRNAs that are targets of the cardiac-expressed miRs, ie, the cardiac RISCome (Figure 2c, upper right quadrant; Online Table I).

Published studies have identified 139 consensus cardiac-expressed miRs (Online Table II). Because the mouse cardiac RISCome should consist of mRNAs that interact with these (and any other) cardiac-expressed miRs, we compared our results with RISC-Seq to the targets predicted for these 139 cardiac-expressed miRs by the TargetScan algorithm. A total of 1171 of 1645 RISC-enriched mRNAs (71.2%) were predicted targets of one or more cardiac-expressed miRs (Figure 2d), compared with only 45.7% of all cardiac-expressed mRNAs. By contrast, only 8 of 75 RISC-depleted mRNAs (10.7%) are predicted cardiac miR targets (Figure 2d; *P=0.0003, Fisher exact test). Thus, our RISC-Seq results are supported by bioinformatic predictions based on established cardiac miR expression signatures.

RISC enrichment scoring ranks transcripts by abundance in RISCome relative to transcriptome (simply: [RISC-mRNA]/[transcript mRNA] for each individual mRNA). Thus, reciprocal regulation of mRNAs in the RISCome and transcriptome, as expected with miR-induced mRNA destabilization, produces a higher score. Because this ranking method does not account for absolute mRNA expression, a transcript expressed at 1 copy per cell and enriched 2-fold in the RISC is scored the same as a transcript having 50 copies per cell with similar RISC enrichment. To determine whether the absolute number of mRNA copies in the RISCome provided critical information about miR targets, we ranked the RISC-associated transcripts according to their absolute abundance and determined the proportion of TargetScan-predicted cardiac miR targets by quintile. RISC-associated mRNA abundance and the likelihood of being a predicted cardiac miR target were associated (*P<0.01, χ² test; Figure 3a), demonstrating that absolute RISC mRNA abundance is a factor in miR targeting. Surprisingly, however, the association between RISC enrichment score and either transcriptome mRNA abundance (Figure 3b) or RISC-associated mRNA abundance (Figure 3c) is inverted, revealing that the most RISC-enriched targets tend to be the rarer mRNA transcripts. This finding is consistent with previous observations that most abundant cardiac mRNAs, encoding myofilament and other constitutively expressed proteins, tend to undergo the least regulation. These results indicate that the analytic approach of relating RISC mRNA abundance to transcript mRNA abundance, ie, RISC enrichment, is superior to just identifying mRNAs in the RISCome.

### Programming of the Cardiac RISCome to Identify Individual miR Targets

A thorough understanding of the pathological roles of dysregulated miRs requires delineating individual in vivo miRNA functional pairs, which has been problematic using bioinformatics predictions and tissue culture experiments. We considered that the cardiomyocyte RISC could be programmed to retain specific miR-targeted mRNAs by cardiomyocyte-specific overexpression of a miR precursor. RISC-Seq comparison of miR-programmed and unprogrammed cardiac RISComes would then identify mRNAs targeted by the programming miR. We tested this notion using RISC-Seq of our miR-133a transgenic mice. miR-133a is a muscle-specific miR essential for cardiac development that is dysregulated in cardiac disease. Importantly, miR-133a transgenic mice have no basal cardiac phenotype to confound miR target analysis by dysregulating cardiac gene expression.

RNA sequencing of the miR-133a–programmed transcriptome and RISCome was performed as described above. Sequencing read depth (1.5 to 2.0×10⁶ mRNA-matching reads per barcoded sample), alignment to the mouse genome, and the identities of detected transcripts were similar for all 16 libraries (1 RNA-seq and 1 RISC-seq library from each of 4 nontransgenic and 4 miR-133a transgenic mice). RISC scoring identified 2149 RISC-enriched mRNAs in miR-133a–programmed hearts, 684 of which had not met criteria for being RISC-enriched in nontransgenic hearts (“new miR-133a targets”), and 209 of which were enriched to a significantly greater extent (>1.3-fold greater RISC score, *P<0.001, χ² test) in the miR-133a transgenic hearts, representing 43 known cardiac mRNAs and 1712 novel mRNAs. These newly identified RISC-enriched cardiac mRNAs are regulated by miR-133a and provide novel targets for the study of cardiac disease.
miR-133a cardiac RISCome (‘hyperenriched miR-133a targets’) (Figure 4a and Online Table III). TargetScan reports only 435 predicted mouse miR-133a targets among 17 315 total mouse transcripts in the database: 310 of these mRNAs are among the ~10 000 mRNAs present in the cardiac transcriptome (~2.8% of all cardiac-expressed mRNAs). The 1645 mRNAs enriched in the normal cardiac RISCome contain 107 (34.5%) of the 310 TargetScan predicted cardiac miR-133a targets, consistent with miR-133a being one of the more highly expressed cardiac miRs.6,25 Finally, 147 (47.4%) of the 310 predicted targets were present in miR-133a-programmed transcriptomes (Figure 4b). Thus, bioinformatics analysis shows a progressive increase in the proportion of predicted miR-133a targets from the cardiac transcriptome to the cardiac RISCome and to the miR-133a-programmed cardiac RISCome.

Enrichment of miR-133a targets in programmed cardiac RISCs and the presence of miR-133a–programmed RISC mRNAs previously shown to be miR-133a targets (Whsc2,25 Hcn4,26 Ccn2,27 Casp9,29 and Ctgf28; only Whsc2 and Ctgf predicted by TargetScan) validated our method of individual miR target analysis by programmed RISC-Seq. However, a number of miR-133a–programmed RISC-enriched mRNAs identified experimentally were not predicted by TargetScan. To determine the reason for this discrepancy, we characterized miR-133a effects on 7 miR-133a–programmed RISC-associated mRNAs using standard luciferase reporter assays.3,10 All 5 miR-133a RISC-enriched mRNA 3’ UTRs were significantly suppressed by miR-133a (Figure 5), including 3 (Fosl2, Pparcg1b, and Sdc3) that were not identified by TargetScan as miR-133a targets. Only 1 of the 2 TargetScan-predicted/RISC-Seq not predicted miR-133a targets was suppressed (Edeml), which was indeed increased 2.5-fold in miR-133a–programmed RISComes, but with a probability value [0.016] that did not meet our selection criteria (Figure 5). Based on these results and the 5 previously validated miR-133a targets, RISC-Seq exhibits a sensitivity of 91% for target identification, versus 45% for TargetScan.

**Discussion**

Here, we describe RISC-Seq, a method that marries RNA sequencing to Ago2 immunoprecipitation for in vivo RISCome profiling and identification of specific miR targets. Our method avoids experimental bias introduced by RNA amplification and takes advantage of superior mRNA-profiling results from RNA sequencing, compared with expression arrays.8 Because our technique works with endogenous Ago2 protein and does not require crosslinking or sample pooling, it is broadly applicable to miR target analysis in genetically and physiologically modeled in vivo models.
The cardiac system is especially well suited for analysis of endogenous miR targets because miR effects are central to embryonic heart formation, reactive cardiac hypertrophy, excitation–contraction coupling, and programmed cardiomyocyte death. miR expression is dynamically regulated in heart disease, and so miRs are being evaluated as diagnostic tools and therapeutic targets. The complement of individual cardiac-expressed mRNAs that are potentially regulated by cardiac-expressed miRs (i.e., the cardiac RISCome) was not previously known. Here, RISC-Seq defined the normal cardiac RISCome consisting of 2000 of 10,000 cardiac-expressed mRNAs. Thus, 20% of cardiac mRNAs are regulated by miRs under normal conditions. RISC-Seq should be equally useful to define RISC-associated mRNAs in pathological conditions, such as cardiac hypertrophy or heart failure, where the miRome and transcriptome are jointly dysregulated.

Our studies show that absolute RISC-associated mRNA abundance is inferior to RISC enrichment in identifying miR targets. Loosely associated, highly abundant transcriptome mRNAs are unavoidably captured during Ago2 immunoprecipitation, explaining the nearly complete representation of cardiac-expressed mRNAs in the RISC complexes. We controlled for nonspecific representation of highly expressed transcripts in Ago2 immunoprecipitates by relating their abundance in the RISCome to that in the transcriptome, setting a threshold level of 2-fold enrichment. We found that lower abundance mRNAs were disproportionately recruited to the RISCome, consistent with recent results demonstrating that the most abundant mRNA transcripts are also subjected to less transcriptional regulation in disease.

To attack the question of individual miR targeting in the heart, we used cardiomyocyte-specific transgenic expression of miR-133a and miR-499 precursors to “program” the cardiomyocyte RISCome. These 2 well-studied miRs have central roles in cardiac development and myofilament identity, respectively, and are regulated in cardiac disease to approximately the levels at which they were overexpressed. In both instances, we identified mRNA targets that were not predicted by TargetScan. These findings demonstrate a significant impact of other determinants of miR-mRNA interactions on RISCome recruitment, suggesting roles for nonseed sequence complementation, the number of complementary miR sequences in a given transcript, and the cumulative effects of other miRs on the same transcript.

Figure 6. Gene ontology (GO) analysis of miR-133a and miR-499 RISCome hyperenriched mRNAs. a, Sequence homology of members of the miR-133 and miR-208/499 families. b, Venn diagram of hyperenriched mRNAs in mir-133a (left) and mir-499 (right) programmed cardiac RISComes with respective GO categories. Overrepresented category compared with distribution of all genes across the transcriptome (see Methods).
We observed this with enriched by the presence of more miR than weaker targets. Furthermore, we think that RISC deprogramming is likely to be of limited utility even in the absence of a confounding phenotype because legitimate miRNA targets of the absent miR will continue to be recruited to the RISC by miR family members having similar sequences (see Figure 6). In addition to providing comprehensive and unbiased analysis of RISComes in pathophysiological context, we believe RISC-Seq in tissues programmed with endogenous miRs will be useful to identify unsuspected mRNA targets having imperfect seed sequence complementarity or extrased sequence binding, which are generally overlooked by bioinformatics. mRNA targets with perfect seed sequence binding, which are generally overlooked by having imperfect seed sequence complementarity or extras-

miR family members having similar sequences (see Figure 6). We observed this with Srf, an experimentally validated miR-133a target that is 2.4-fold RISC-enriched in miR-133a–programmed hearts, but did not meet our criteria for statistical significance ($P<0.0001$), and with Sox6, a validated miR-499 target that is already 5.1-fold RISC-enriched in nontransgenic hearts and only increases a further 1.2-fold (below our threshold for 1.3-fold change) in miR-499 programed hearts. Thus, RISC programming with endogeo

nously expressed miRs may have the greatest proportional impact on the less obvious, and most difficult to define, mRNA targets. Perhaps in combination with RNA footprinting to define miR-mRNA binding domains and proteomics studies to elucidate the biology of miR effects, RISC-Seq will be especially useful to define the in vivo targets of synthetic “designer” miRs and miR-mimetics undergoing evaluation as therapeutics.

Sources of Funding

This work was supported by NIH grants RC2 HL102222, R01 HL59888, and U1L RR024992.

Disclosures

None.

References


**Novelty and Significance**

**What Is Known?**

- MicroRNAs (miRs) are important regulators of gene expression, play roles in cardiovascular development and disease, and may serve as novel cardiovascular therapeutics.
- Although a single miR can potentially suppress hundreds of mRNAs, comprehensive identification of in vivo targets is lacking.

**What New Information Does This Article Contribute?**

- Methods were developed for deep sequencing of mRNAs targeted by miRs into RNA-induced silencing complexes (RISCs). The method avoids bias introduced by RNA amplification, cross-linking of RNA to RISCs, and use of expression arrays.
- Identification of mRNAs enriched in wild-type mouse cardiac RISCs was performed by comparing expression levels in the RISCome (RISC-associated RNAs) to that in the transcriptome (poly(A)-selected RNA).
- Programming cardiomyocyte RISCs via transgenic overexpression of miRs reveals specific profiles of RISC-enriched miRNAs in intact mouse hearts, and thus the in vivo targets of miRs in cardiomyocytes.

MicroRNAs (miRs) are key regulators of mRNA translation in health and disease. Although bioinformatic predictions suggest that a single miR may target hundreds of mRNAs, the number of experimentally verified targets of miRs is low. To enable comprehensive, unbiased examination of miR targets, we have performed deep RNA sequencing of cardiac transcriptomes in parallel with cardiac RNA-induced silencing complex (RISC)-associated RNAs (the RISCome), called RISC sequencing. We developed methods that did not require crosslinking of RNAs to RISCs or amplification of mRNA before sequencing, making it possible to rapidly perform RISC sequencing from intact tissue while avoiding amplification bias. Comparison of RISCome with transcriptome expression defined the degree of RISC enrichment for each mRNA. The majority of the mRNAs enriched in wild-type cardiac RISComes compared with transcriptomes were bioinformatically predicted to be targets of at least 1 of 139 cardiac-expressed miRs. Programming cardiomyocyte RISCs via transgenic overexpression of miRs revealed specific profiles of RISC-enriched miRNAs in intact mouse hearts, and thus the in vivo targets of miRs in cardiomyocytes.
RISC RNA Sequencing for Context-Specific Identification of In Vivo MicroRNA Targets
Scot J. Matkovich, Derek J. Van Booven, William H. Eschenbacher and Gerald W. Dorn II

Circ Res. published online October 28, 2010;
Circulation Research is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2010 American Heart Association, Inc. All rights reserved.
Print ISSN: 0009-7330. Online ISSN: 1524-4571

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://circres.ahajournals.org/content/early/2010/10/28/CIRCRESAHA.110.233528.citation

Data Supplement (unedited) at:
http://circres.ahajournals.org/content/suppl/2010/10/28/CIRCRESAHA.110.233528.DC1

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in Circulation Research can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:
http://www.lww.com/reprints

Subscriptions: Information about subscribing to Circulation Research is online at:
http://circres.ahajournals.org//subscriptions/
SUPPLEMENTAL METHODS

Preparation and quantification of total myocardial RNA
Cardiac miR-133a transgenic mice were described previously 1. miR-499 transgenic mice were generated similarly by cloning a 530 bp fragment flanking mouse miR-499 (chromosome 2, intron 19 of Myh7b) into the Myh6 promoter construct. Total cardiac RNA was isolated from flash-frozen ventricular tissue using Trizol (Invitrogen) and quantified on a UV spectrometer.

Reverse transcription and preparation of cDNA
Preparation of cDNA fragments was modified from previously described protocols 1, 2. For poly(A)+ RNA, 4 μg of total cardiac RNA was twice oligo(dT) selected using the Dynabead mRNA purification system (Invitrogen). Two hundred ng of cardiac mRNA was fragmented to ~200 nt by heating to 94°C for 2.5 min in 40 mmol/L Tris acetate pH 8.2, 100 mmol/L potassium acetate, 30 mmol/L magnesium acetate, and immediately chilled on ice. After purification on Ambion NucAway columns, 100 ng of fragmented cardiac mRNA was reverse-transcribed using SuperScript III (Invitrogen) with random hexamers as per the manufacturer’s directions (50 min, 50°C), followed by second-strand cDNA synthesis for 2 h, 16°C, in 20 mmol/L Tris.HCl pH 6.9, 90 mmol/L KCl, 4.6 mmol/L MgCl2, 150 μmol/L β-NAD+, 10 mmol/L (NH4)2SO4, 0.067 μL/μL E. coli DNA ligase, 0.27 μL/μL E. coli DNA polymerase I, and 0.013 μL/μL E. coli RNase H (New England Biolabs).

Integrated reverse-transcription and amplification of low-input RNA
RNA amplification prior to fragmentation and sequencing library construction was performed as per Mitreva and Mardis 3. Briefly, polyA-tailed RNA was selected from total RNA using a primer recognizing 5’ mRNA ends and one containing an oligo(dT) moiety. Following reverse transcription, a third primer recognizing common elements in the first two primers was used to amplify cDNAs and to add biotin tags. Optimization of amplification cycles was performed as described 3 and biotin tags were removed following restriction enzyme digestion and removal on streptavidin beads.

Immunoprecipitation of mouse cardiac Ago2-associated RNA and preparation of cDNA
Conditions for Ago2 immunoprecipitation were adapted from Karginov et al. 4. Frozen mouse heart bases were homogenized in 500 μL ice-cold 50 mmol/L Tris.HCl, 5 mmol/L EDTA, 5 mmol/L EGTA, pH 7.5, with Roche Complete protease inhibitors. Yeast tRNA (Invitrogen) and SUPERnase-IN (Ambion) were added to final concentrations of 1 mg/mL and 1 U/μL, respectively, and unbroken cellular material was removed at 100g, 5’, 4°C. Nonidet P-40 was added to a final concentration of 0.5% (w/v) to solubilize proteins (15’, 4°C, rotating) and insoluble material was removed at 10000g, 15’, 4°C. The supernatant was added to 50 μL protein G-coupled Dynabeads (Invitrogen), to which 5 μg anti-mouse Ago2 monoclonal antibody (Wako Pure, clone #2D4, lot PEM0820) had been previously bound according to the Dynabead protocol. Following 1 h rotational incubation at 4°C, the beads were washed 3x with Dynabead washing buffer, transferring the suspension to a fresh tube for the last wash. Beads were pelleted from the remaining suspension, the supernatant was removed, and 500 μL Trizol was added to the immunoprecipitated material to extract RNA. Ago2 immunoprecipitate-associated RNA was fragmented in acetate buffer without poly-A+ selection, purified on NucAway columns, and one half was used for cDNA synthesis using the protocol described above.

Construction of DNA-barcoded short-read libraries for Illumina sequencing
Detailed methods for preparation of Illumina sequencing libraries from mouse cardiac RNA were recently described 1. Briefly, cDNAs were end-repaired using the End-It End-Rearray kit (Epicentre Biotechnologies, #ER0270) and 3’ A-overhangs added using 3’-5’ exo- Klenow polymerase (New England Biolabs #M0212) and 0.2 mmol/L dATP. Illumina adapters with T-overhangs and customized to include three nt ‘barcodes’ were ligated to the cDNA at 10:1 molar excess using the Promega LigaFast kit.
Different barcoded adapters were ligated to individual mouse heart cDNAs. Following column-purification (Qiagen) to remove excess unligated adapter, DNA in the 200-400 bp range was isolated via gel purification (Qiagen) on 2% low-melting agarose. One-third of the gel-purified material was amplified with 12 cycles (total myocardial cDNA) or 16 cycles (RISC-immunoprecipitated cDNA) of Phusion polymerase (New England Biolabs #F531)-mediated PCR (10 sec 98 C, 30 sec 65 C, 30 sec 72 C cycles, followed by final 5 min 72 C), using oligonucleotides complementary to Illumina sequencing adapters. The final, amplified libraries were again column-purified and quantified using PicoGreen (Quant-It, Invitrogen).

Four barcoded libraries were combined in equimolar (10 nmol/L) amounts and diluted to 6 pmol/L for cluster formation on a single Illumina Genome Analyzer II flowcell lane, followed by single-end sequencing. Basecalling of DNA clusters was performed using Illumina’s processing pipeline software (version 1.5) and 36-nt sequences, with quality scores, were obtained in Illumina’s SCARF text format. Barcoded sequences were deconvoluted and sorted as previously described 1. After barcode removal, the 32 base mRNA sequence reads were mapped to transcripts annotated in NCBI release 37 of the mouse genome using the publicly available packages Bowtie (release 0.12.3) (http://bowtie-bio.sourceforge.net/index.shtml) 5, TopHat (release 1.0.13) (http://tophat.cbcb.umd.edu/) 6, and Cufflinks (release 0.8.1) (http://cufflinks.cbcb.umd.edu/) 7, as previously described 1. Cufflinks was used with gene annotation files to calculate overall gene expression in terms of Fragments Per Kb of exon per Million mapped reads (FPKM) 7. When performing single-end Illumina sequencing, as we have done here, this parameter is equivalent to RPKM (Reads Per Kb of exon per Million mapped reads), in which an RPKM of 3 corresponds to 1 copy/cell in cardiac samples 1, 2. We used the default options supplied with these software packages in our analyses and analyzed only those RNA elements that had expression signals in at least 2 of 4 biological replicates.

**Computation of RISC abundance and enrichment scores**

Because they are not poly-A+ selected, RISComE sequencing libraries contained a higher proportion of ribosomal RNA (32-42%) than poly-A+ selected transcriptome libraries (3-5%). To permit comparison between the transcriptome and RISComE, measures of RNA abundance expressed as Fragments Per Kb of exon per Million mapped reads (FPKM 7) were adjusted for the proportion of mRNA. RISC enrichment scores to assess RISC enrichment or depletion were defined as (RISComE adjusted FPKM/transcriptome adjusted FPKM) ≥2.0 or ≤0.5, P<0.0001, FDR<0.01. Scores <1 were plotted as -1/score for clarity.

**Luciferase reporter constructs**

Full (if ≤1 kb) or partial (regions selected to contain miR-133a binding ‘seeds’) 3’ untranslated regions of candidate miR-133a target genes were cloned into the dual-luciferase reporter vector psiCheck2 (Promega). The miR-133a genomic precursor consisted of a 735-bp fragment flanking the mouse miR-133a-1 locus on chromosome 18 (519-bp 5’ upstream sequence, 21-nt miR-133a, and 195-bp 3’ sequence). 100 ng of psiCheck2 construct, together with 100 ng of pcDNA3.1 or pcDNA3.1+miR-133a genomic precursor was transfected into HEK293 cells on 24-well plates using Fugene HD (Promega). After 48 h, duplicate determinations of Renilla and firefly luciferase activities were performed after harvesting cells into 120 μL Glo Lysis Buffer (Promega). 30 μL of each lysate was analyzed using Dual-Glo Luciferase reagents (Promega) on a SpectraMax M5e platereader (Molecular Devices). Renilla 3’UTR-coupled luciferase activity was normalized to constitutive firefly luciferase activity for each well.

**Statistical analysis**

Gene symbols and FPKM values were imported into Partek Genomics Suite v6.5 (Partek, St Louis, MO) for comparison of RISComE and transcriptome expression values, computation of P-values and false discovery rates, and preparation of P-value vs fold-change (volcano) plots. GraphPad Prism was used for chi-square tests. Luciferase activities were compared using Student’s t-test (GraphPad
Prism). Gene Ontology over-representation was performed using a hypergeometric test with Benjamini & Hochberg false discovery rate correction. Significant differences were defined as P<0.05 unless otherwise described.

REFERENCES

Supplemental Table I is supplied as an Excel datasheet.

**Supplemental Table I.** *RISC-enrichment in nontransgenic mouse hearts.* FPKM (Expression) values, fold-change between RISC-Seq and RNA-Seq, and corresponding p-values are shown for each gene. RISC-enriched genes were defined as those upregulated by at least 2-fold in the RISCome vs the transcriptome, P<0.0001.

Supplemental Table II is supplied as an Excel datasheet.

**Supplemental Table II.** *Cardiac-expressed miRs.* 139 miRs expressed in myocardium.

Supplemental Table III is supplied as an Excel datasheet.

**Supplemental Table III.** *Enrichment of mRNAs in miR-133a transgenic vs nontransgenic RISComes.* First tab: 209 mRNAs hyper-enriched between the miR-133a and nontransgenic RISCome. Second tab: 684 mRNAs enriched in the miR-133a, but not the nontransgenic RISCome.

Supplemental Table IV is supplied as an Excel datasheet.

**Supplemental Table IV.** *Enrichment of mRNAs in miR-499 transgenic vs nontransgenic RISComes.* First tab: 81 mRNAs hyper-enriched between the miR-133a and nontransgenic RISCome. Second tab: 334 mRNAs enriched in the miR-133a, but not the nontransgenic RISCome.
**Supplemental Figure I.** RNA-sequencing count data $\log_2(FPKM+1)$ from the two different unamplified mouse cardiac mRNA samples.
Supplemental Figure II. Heatmap of pathological mRNA expression in Gq transgenic mice; left lanes, nonamplified (Non), right lanes, amplified (Ampl). Blue indicates low expression, red indicates high expression.