microRNA-seq of cartilage reveals an over-abundance of miR-140-3p which contains functional isomiRs

Steven Woods[1], Sarah Charlton[2], Kat Cheung[2], Yao Hao[2,3], Jamie Soul[2], Louise N Reynard[2], Natalie Crowe[4], Tracey E. Swingler[4], Andrew J. Skelton[2], Katarzyna A. Piróg[2], Colin G. Miles[2], Dimitra Tsompani[2], Robert M. Jackson[2], Tamas Dalmay[4], Ian M. Clark[4], Matt J. Barter[2], David A. Young[2]*

[1]Division of Cell Matrix Biology and Regenerative Medicine, Faculty of Biology Medicine and Health, University of Manchester, Michael Smith Building, Oxford Road, Manchester, M13 9PT, UK
[2]Skeletal Research Group, Biosciences Institute, Newcastle University, Central Parkway, Newcastle upon Tyne, NE1 3BZ, UK
[3]Orthopedics Department, First Hospital of Shanxi Medical University, Yingze District, Taiyuan, 030000, China
[4]School of Biological Sciences, University of East Anglia, Norwich, NR4 7TJ, UK
*To whom correspondence should be addressed. Tel: +44 191 2418831; FAX: Not applicable; Email: david.young@ncl.ac.uk

ABSTRACT
MiR-140 is selectively expressed in cartilage. Deletion of the entire miR-140 locus in mice results in growth retardation and early-onset osteoarthritis-like pathology, however the relative contribution of miR-140-5p or miR-140-3p to the phenotype remains to be determined. An unbiased small RNA sequencing approach identified miR-140-3p as significantly more abundant (>10-fold) than miR-140-5p in human cartilage. Analysis of these data identified multiple miR-140-3p isomiRs differing from the miRBase annotation at both the 5 ‘ and 3 ‘ end, with >99% having one of two seed sequences (5 ‘ bases 2-8). Canonical (miR-140-3p.2) and shifted (miR-140-3p.1) seed isomiRs were overexpressed in chondrocytes and transcriptomics performed to identify targets. miR-140-3p.1 and miR-140-3p.2 significantly down-regulated 694 and 238 genes respectively, of which only 162 genes were commonly down-regulated. IsomiR targets were validated using 3 ‘UTR luciferase assays. miR-140-3p.1 targets were enriched within up-regulated genes in rib chondrocytes of *Mir140*-null mice and within down-regulated genes during human chondrogenesis. Finally, through imputing the expression of miR-140 from the expression of the host gene *WWP2* in 124 previously published datasets, an inverse correlation with miR-140-3p.1 predicted targets was identified. Together these data suggest the novel seed containing isomiR miR-140-3p.1 is more functional than original consensus miR-140-3p seed containing isomiR.

INTRODUCTION

MicroRNAs (miRNAs) are small non-coding RNAs that regulate gene expression. Mature miRNAs are processed from primary transcripts containing stem loops by cleavage mediated by Drosha and then Dicer proteins (1). The 5′ and 3′ sides of the stem loop can both give rise to mature miRNAs termed ′-5p′ and ′-3p′ respectively (2). These mature miRNAs are then loaded into the RNA-induced silencing complex (RISC) to mediate either mRNA degradation or translation inhibition of target mRNA through binding of the miRNA seed sequence (5′ bases 2-8) to the target 3′UTR (3). The level of complementarity between the seed and target mRNA is important for target repression. Binding to positions 2-7 of the miRNA may indicate target repression, however either an adenine binding to position 1 of the miRNA (7a1 targets), a match at position 8 (7m8 targets) or both (8mer targets) can improve target repression (4).

In recent years, small RNA (sRNA) sequencing (sRNA-seq) has identified additional miRNAs that do not perfectly align to the annotated mature miRNA, known as isomiRs (5). IsomiRs may be shorter or longer at the 5′ and/or 3′ end of the mature miRNA, and are not always templated to the genomic DNA. Variants at the 3′ end are quite common, although not thought to alter the miRNA target repertoire (6). Variations at the 5′ end are less common, but are of greater importance due to altered seed sequence, target repertoire and potentially the function of miRNAs.

The most studied miRNA in cartilage is miR-140, which produces miR-140-5p and miR-140-3p (7,8). Deletion of *Mir140* in mice leads to a skeletal phenotype including an OA-like disease. However, the contribution of the lack of miR-140-5p or miR-140-3p to the phenotype remains largely undetermined. Barter *et al*. pay particular attention to the role of miR-140-5p and suggests there are many genes under its control during chondrogenesis (9). miR-140-3p is less well studied and only a small number of -3p targets have been identified, which so far appear to have little relevance to cartilage and OA. Intriguingly, using small RNA sequencing (sRNA-seq) of human cartilage RNA we identified that miR-140-3p was high in abundance (>10-fold) compared to miR-140-5p (10).

Although there have been a number of studies to elucidate the role of miRNAs in cartilage and chondrogenesis, the role of isomiRs has been largely ignored. Here we show miR-140-3p isomiRs are abundantly expressed in cartilage and that these isomiRs are present in RISC. Using over expression followed by transcriptomic analysis we show two miR-140-3p isomiRs (miR-140-3p.1 and miR-140-3p.2) have largely discrete target repertoires. We validated a number of common and discrete targets for each isomiR using a luciferase reporter system. We present evidence that miR-140-3p.1, which is not currently annotated in miRBase, may play roles in human chondrogenesis, mouse cartilage and in multiple other skeletal tissues.

MATERIAL AND METHODS

Analysis of cartilage sRNA-seq

sRNA-seq was analysed as previously described (10). 5' isomiRs were defined using the following criteria: 1) loss or gain of 1 or more nucleotides, using the mature miRBase (11) sequences as a reference; 2) a read count >100, and 3) a read count > 5% of the mature miRBase reference sequence for that miRNA.

Human articular chondrocyte isolation, culture and transfection

Human articular chondrocyte isolation from knee cartilage was performed as previously described (9). Tissue was donated by patients with diagnosed osteoarthritis and undergoing joint replacement surgery, with informed consent and ethics committee approval. Briefly, macroscopically normal cartilage was removed from the subchondral bone and dissected into ~1mm pieces using scalpel and forceps. Enzymatic digestion was performed using trypsin and then collagenase overnight at 37°C. Chondrocytes were then grown to confluence and plated into 6 well plates for miRNA/isomiR mimic transfection. 100nM miRNA mimic and control were transfected into HAC using DharmaFECT 1 transfection reagent (Dharmacon, Horizon Discovery, Cambridge, UK) according to the manufacturer's protocol and essentially as previously described (9). 48h post-transfection HAC were lysed and RNA harvested using Qiagen miRNeasy kit (Qiagen, Crawley, UK). Custom miRNA mimics with each isomiR sequence were purchased from Dharmacon.

Microarray and identification of targets

Microarray was performed by Cardiff University, Central Biotechnology Services, using Illumina whole-genome expression array Human HT-12 V4 (Illumina, Saffron Walden, U.K.). Normalization of the quantified signals (background corrected) was performed using the global Lowess regression algorithm. Expression analysis was performed in R/bioconductor using the limma package (12). PCA and heat map were generated using ClustVis (13). PCA analysis was performed on all unique genes detected by the microarray analysis. Heat map is based on the top 650 significant genes for each comparison (miC vs. mR-140-3p.1 or miC vs. miR-140-3p.2), which after removal of duplicates gave 882 genes. TargetScanHuman (version 7.2 (14)) was used to predict targets of miR-140-3p.1 and miR-140-3p.2. In order to demonstrate discrete functions we investigated unique targets for each isomiR (3p.2 and 5p targets were removed from 3p.1 analysis; 3p.1 and 5p targets were removed from 3p.1 analysis; 3p.1 and 3p.2 targets were removed from 5p analysis). Genes were selected for luciferase validation based on both altered expression in miR-140-3p.1 or miR-140-3p.2 transfected cells versus control and on expression in miR-140-3p.1 transfected cells versus expression in miR-140-3p.2 transfected cells.

Target enrichment and pathway analysis

Sylamer analysis was performed on ordered gene lists from most downregulated to most upregulated (including genes whose expression did not significantly change) (15). Target enrichment (%targets) was calculated by dividing the total number of predicted targets (TargetScan 7.2) that significantly increased or decreased by the total number of genes that significantly increased or decreased, multiplied by 100. Cumulative fraction plots were generated using ordered gene list from down-regulated to most up-regulated (regardless of significance) and predicted targets from TargetScan 7.2. Pathway analysis was performed using Database for Annotation, Visualization and Integrated Discovery (DAVID) v6.7(16).

Q-RT-PCR for miR-140-3p.1 and miR-140-3p.2

Custom miR-140-3p.1 and miR-140-3p.2 assays (Exiqon) were used to detect expression of miR-140-3p.1 and miR-140-3p.2. Assays were validated using spike-in of miR-140-3p.1 and miR-140-3p.2 mimic (Dharmacon).

3´UTR luciferase reporter construction and assay

In-Fusion® cloning (Clontech, Takara Bio Europe SAS, Saint-Germain-en-Laye, France) of selected 3'UTRs into the pmirGLO vector (Promega, Southampton, UK) was used to generate 3'UTR luciferase reporters essentially as previously described (9). SW1353 chondrosarcoma cells were seeded and cultured to reach ~50% confluence after 24 h (9). miRNA (100 nM) were transfected using DharmaFECT 1 transfection reagent, reporter plasmids (500 ng/ml) were transfected using FugeneHD (Promega). 24h after transfection luciferase levels determined using Promega dual luciferase assay and GloMax plate reader (Promega).

Generation of miR-140$^{-/-}$ mice

All animal experiments were performed under licenses granted from the Home Office (United Kingdom) in accordance with the guidelines and regulations for the care and use of laboratory animals outlined by the Animals (Scientific Procedures) Act 1986 according to Directive 2010/63/EU of the European Parliament, and conducted according to protocols approved by the Animal Ethics Committee of Newcastle University and the Home Office, United Kingdom. Post-generation breeding and subsequent phenotyping were performed under licences PPL60/4525 and P8A8B649A. CRISPR/Cas9 guide RNAs (crRNA) were designed using CHOPCHOP. crRNA linked with TRACR (sgRNA) were amplified by PCR with a pLKO vector (Addgene_52628) as template, using the primer T7 TRACR R (Supplementary Table 10) and a 5' PCR primer that included a T7 sequence and the crRNA. This was converted to RNA using the MEGAshortscript T7 kit (Fisher Scientific). sgRNA (50ng/ml each) were mixed with recombinant Cas9 (ToolGen, CamBioScience Limited, Cambridge, UK) and injected into the cytoplasm of donor mouse zygotes and transferred into recipient foster mothers, all essentially as previously described (17-19). The mixed C57BL/6 and CBA/ca F_0 mice were backcrossed onto C57BL/6J and heterozygous animals crossed three times to eventually generate wild-type and null lines used in purification of rib chondrocytes.

Genotype was confirmed by ear-notch PCR and Sanger sequencing (Supplementary Figure 3).

Rib chondrocyte isolation and RNA-seq

Primary mouse costal chondrocytes were isolated from 7-day-old miR-140$^{-/-}$ and WT mice using collagenase digestion, essentially as previously described (20). Total RNA, including miRNA was isolated using the miRVana miRNA Isolation Kit (with phenol)(Fisher Scientific). Sequencing libraries were prepared from 500 ng of purified total RNA using the Illumina TruSeq Stranded mRNA sample preparation kit according to the manufacturer's protocol, and sequenced on Illumina NextSeq500. Each sample provided >12 million single-end 75-bp sequencing reads. Sequenced reads were mapped to the mm10 transcriptome using Salmon (21). Batch effects were estimated using RUVseq (22) and incorporated into the differential expression analysis performed using DESeq2 (23). PCA and heat map were generated using ClustVis (13).

SkeletalVis analysis

Gene expression responses within SkeletalVis were filtered for human and mouse data sets and where *WWP2* expression significantly changed (adjusted p-value is <0.05), leaving 124 experimental comparisons. The change in WWP2 expression was plotted against the average fold change of predicted (TargetScan7.2) miRNA targets for each experimental comparison. Correlation (R^2) and regression analysis for all 124 experimental comparisons was calculated using the data analysis add-in in Microsoft Excel. Target enrichment (%targets) was calculated by dividing the total number of predicted targets (TargetScan 7.2) that significantly increased or decreased by the total number of genes that significantly increased or decreased, multiplied by 100, for each of the 124 experimental comparisons. miR-140-5p, miR-140-3p.1 and miR-140-3p.2 target enrichment within up, no change or down-regulated genes was then plotted against log2 FC for *WWP2*. Cumulative mean enrichment (from studies where *WWP2* is most up- or down-regulated compared to studies where *WWP2* changes least, were plotted to demonstrate trend. Correlation between miR-140-5p, miR-140-3p.1 and miR-140-3p.2 target enrichment was plotted by first calculating the relative enrichment in up vs down-regulated genes for each experimental comparison (enrichment in up/ enrichment in down). This relative enrichment for each of miR-140-5p, miR-140-3p.1 and miR-140-3p.2 was plotted against each other. Experimental comparisons where *WWP2* decreased or increased were shown in different colours as indicated in figure legend. A list of randomly generated genes was analysed alongside miR-140-5p, miR-140-3p.1 and miR-140-3p.2.

RESULTS

RNA-Seq of articular cartilage identifies an overabundance of miR-140-3p and miR-140-3p isomiRs

We performed sRNA-seq of human cartilage RNA and identified novel cartilage specific miRNAs (10). Our cartilage sRNA-seq aligned to 990 miRNAs annotated in miRbase (11), 704 of which contained at least one additional type of potential isomiR (Table S1). As 5′ isomiRs are most likely to alter function, we focused further sRNA-seq analysis on miRNAs with 5'isomiRs. Using thresholds of >100 reads and >5% of total reads for that miRNA, we identified 5′isomiR sequences for 29 miRNAs, 26 of which have a single addition or deletion at the 5′ nucleotide, while miR-3074-5p, miR-455-3p and let-7b-3p all have two additional 5′ nucleotides (Supplementary figure 1). Although the isomiR for miR-1246 is just under our read threshold (read count of 99), it appears to have four additional 5′ nucleotides compared to the miRBase annotation (Supplementary figure 1).

MiR-140-3p and its isomiRs account for more than half of all sequencing reads in our cartilage sRNA-seq (Figure 1A). IsomiRs can be sub-divided into several categories, essentially either templated or non-templated with either 5' or 3' modifications (24). miR-140-3p isomiRs are predominantly 3' additions and mixed type isomiRs, with only 5% aligning to the original miRBase annotation (Figure 1B). More than 99% of all miR-140-3p isomiRs result in one of two seed sequences; ACCACAG as annotated in miRBase, and CCACAGG, which is shifted by -1 nucleotide (Figure 1C). Twenty-five percent of reads for miR-140-3p were perfectly templated to DNA (Figure 1D), the remainder contained one or more non-templated nucleotides, predominantly non-templated 3' adenine additions (Figure 1E), which were observed in sequencing reads containing both of the seed sequences (Figure 1F and G). The most detected isomiR with each seed are termed miR-140-3p.1 (ACCACAGGGTAGAACCACGGAC, seed CCACAGG) and miR-140-3p.2 (TACCACAGGGTAGAACCACGGA seed: ACCACAG), respectively (Figure 1H). Expression of miR-140-3p.1 and miR-140-3p.2 in cartilage was validated using qRT-PCR with isomiR selective assays (Figure 1I).

Analysis of other published sRNA-seq data identified the presence of miR-140-3p.1 and miR-140-3p.2 in other human tissue types, although expressed at a lower level than in cartilage (Supplementary figure 2) (25-27). As miRNA bound to Argonaute (AGO) indicates the ability to repress targets (28), we analysed sRNA-seq data following immunoprecipitation of AGO proteins (29). Indeed both miR-140-3p.1 and miR-140-3p.2 were present, indicating each can be loaded into RNA-induced silencing complex (RISC; Supplementary figure 2), and have the potential to target transcripts.

miR-140-3p.1 and miR-140-3p.2 are predicted to have distinct targets

As miRNAs generally target mRNAs through interaction of their seed sequence (nucleotides 2-7), the seed sequences for miR-140-3p.1 and miR-140-3p.2 differ (CCACAGG and

ACCACAG, respectively; Figure 2A). Thus each miR-140-3p isomiR has different preferences for seed binding sites (Figure 2A). Indeed, target prediction analysis for miR-140-3p.1 and miR-140-3p.2 by TargetScan 7.2 (14) (conserved) identifies only 133 shared targets, a small number compared to the total number of predicted targets (665 and 496 for miR-140-3p.1 and miR-140-3p.2, respectively, Figure 2B, Table S2). The majority of these targets also differ from predicted miR-140-5p targets (Figure 2B, Table S2).

Identification and validation of miR-140-3p.1 and miR-140-3p.2 targets in human chondrocytes

Next, we used overexpression of miR-140-3p.1 and miR-140-3p.2 in primary human chondrocytes followed by genome-wide transcriptomics to experimentally identify similarities and differences in their target repertoire. Overexpression of miR-140-3p.1 and miR-140-3p.2 was validated using isomiR-specific qRT-PCR (Figure 3A). Principal component analysis (PCA) and heat map illustrated clustering of biological replicates and differences between miR-140-3p.1 and miR-140-3p.2 in transfected chondrocytes (Figure 3B and C). 693 and 237 genes were significantly (adj.P.Val<0.05) down-regulated by miR-140-3p.1 and miR-140-3p.2 respectively, however only 161 genes were commonly down-regulated by both isomiRs (compared to control;Figure 4D, Table S3, Table S4). Pathway analysis of repressed genes identified enrichment terms associated with 'phosphoprotein' for miR-140-3p.1 (Table S5). Direct comparison of miR-140-3p.1 transfected with miR-140-3p.2 transfected cells revealed 237 genes were significantly different between the two isomiRs (Table S3, Table S4). Pathway analysis of the 237 genes differentially expressed between the two isomiRs identified an enrichment of triple helix collagen genes (*COL2A1, COL4A1, COL6A3, COL11A1, COL11A2*; p=0.0004, adjp=0.066; Table S5), indicating the sequence difference between the two isomiRs is functionally important for chondrocytes/cartilage.

Sylamer, a non-bias motif analysis tool, was used to investigate miRNA binding site enrichment within downregulated 3'UTRs (15). Transfection of either miR-140-3p.1 or miR-140-3p.2, showed specific enrichment for 8mer, 7m8, 7a1 and 6mer seed binding sites of the transfected isomiR. Interestingly, a greater enrichment was observed for miR-140-3p.1 (non-canonical seed) binding sites than miR-140-3p.2 (canonical seed) binding sites (Figure 3E). Predicted targets (identified using TargetScan 7.2) that are unique to either miR-140-3p.1 or miR-140-3p.2 were downregulated following overexpression of either miR-140-3p.1 or miR-140-3p.2 respectively, but not downregulated by the reciprocal isomiR (Figure 3F), suggesting both isomiRs can repress specific transcripts in human chondrocytes.

Unique and common targets were then selected for 3′UTR luciferase validation (Figure 3G-I). As expected *LIPA*, *RCL1*, *COL4A1* and *SLC35D2* were validated as miR-140-3p.1 targets; *NUDT21*, *REEP5* and *SDC4* were validated as miR-140-3p.2 targets and *PAX2* validated as a target of both miR-140-3p.1 and miR-140-3p.2. Interestingly, the *REEP5* 3′UTR luciferase constructs was also repressed by miR-140-3p.1, although not a predicted target of miR-140-

3p.1. Conversely, *DAAM1* and *ALDH1A3* are predicted targets of both miR-140-3p.1 and miR-140-3p.2, however significant repression was only observed for miR-140-3p.1 and not for miR-140-3p.2. Contrary to the repression role of miRNAs in gene regulation, miR-140-3p.1 increased luciferase for the *MARCKSL1*, *KIAA1199* and *MAPILC3A* 3'UTR reporters.

miR-140-3p isomiRs are functional in mouse cartilage

Analysis of mouse sRNA-Seq data revealed that both miR-140-3p.1 and miR-140-3p.2 are present and more abundant than miR-140-5p (30) (Figure 4A-C), indicating miR-140-3p isomiRs are conserved across species. We and others have generated *Mir140* null mice, which lack all miRNAs and isomiRs produced from the *Mir140* locus (Supplementary figure 3). Whole genome RNAseq data of rib cartilage from these mice identified 1894 and 1179 genes that were significantly up and down-regulated respectively (Table S6, Table S7). Pathway analysis of upregulated genes following knock out (KO) of *Mir140* identified terms associated with 'phosphoprotein' (Table S8), which is consistent with pathway analysis of genes decreased following miR-140-3p.1 overexpression in human chondrocytes (Table S5). miR-140-5p, miR-140-3p.1 and miR-140-3p.2 predicted targets (identified using TargetScan 7.2) were enriched within upregulated genes (Figure 4D), and their average expression increased following KO of the *Mir140* locus (Figure 4E), indicating a loss of target repression. Furthermore the cumulative fraction plot indicates the fraction of targets that decrease following KO of the *Mir140* locus is lower than expected (Figure 4F). Non-biased Sylamer analysis also identified an enrichment of miR-140-5p, miR-140-3p.1 and miR-140-3p.2 7m8 seed binding sites within upregulated genes; miR-140-5p displayed the greatest enrichment, followed by miR-140-3p.1 and then miR-140-3p.2 (Figure 4G).

More than 99% of miRNAs produced from the murine *Mir140* locus share seed binding sites with either miR-140-5p, miR-140-3p.1 or miR-140-3p.2; of these miR-140-5p has the lowest read count, yet has the largest contribution to target repression (Figure 4). We used Sylamer to investigate the possibility that other lowly expressed sequences produced from the *Mir140* stem loop may contribute to target repression (Supplementary figure 4A). This analysis indicates miR-140-5p followed by miR-140-3p.1 have the largest contribution to target repression in mir140 KO mice (Supplementary figure 4B), and that there are no lowly expressed *Mir140*-derived miRNAs that have a major contribution to target repression in mice (Supplementary figure 4B).

miR-140-3p.1 targets are enriched within down-regulated genes during MSC chondrogenesis

miR-140 is highly up-regulated during mesenchymal stem cell (MSC) chondrogenesis, with miR-140-5p predicted targets highly enriched within the down-regulated genes (9). Here we show through re-analysis of this transcriptomic dataset (Figure 5A) that expression of predicted targets of both miR-140-3p.1 and miR-140-3p.2 are decreased during chondrogenesis. Similar to miR-140-5p predicted targets, the average expression of miR-

140-3p.1 and miR-140-3p.2 predicted targets decreased during chondrogenesis, with unique miR-140-3p.1 targets being more repressed than unique miR-140-3p.2 predicted targets (p=0.07) (Figure 5B). Sylamer analysis indicated enrichment for miR-140-5p and miR-140-3p.1, but minimal enrichment for miR-140-3p.2 (Figure 5C). Consistent with *Mir140* null mouse data, the enrichment of miR-140-5p was greater than for both miR-140-3p.1 and miR-140-3p.2 in genes whose expression decreased during human MSC chondrogenesis (Figure 5C).

miR-140-5p and miR-140-3p.1 predicted targets inversely correlate with *WWP2* in multiple skeletal tissues

Having established miR-140-5p, miR-140-3p.1 and miR-140-3p.2 targets are regulated during human MSC chondrogenesis and in mouse rib chondrocyte development, we next looked for evidence of an inverse correlation between miR-140 and target gene expression in other skeletal tissues. Whole genome transcriptomic datasets of skeletal tissues following various perturbations have been collated within SkeletalVis, a data portal for cross-species skeletal transcriptomics data (31). Although the majority of these studies do not directly assess miRNA expression, miRNAs that are located in the introns of protein coding genes, including miR-140, and are frequently co-regulated with their host gene (32). A number of studies have utilised host-gene expression to predict the gene targets of intronic miRNAs through a consistent negative correlation in expression (33,34). miR-140 is located in intron 16 of *WWP2*, with expression of the miR and an abundant isoform of *WWP2* (transcript variant 2) controlled by a common promoter (35). *WWP2* was therefore used as a surrogate for miR-140 expression.

Of the 779 skeletal gene expression responses within SkeletalVis, 124 contained a significant alteration (adjusted p-value <0.05) in *WWP2* expression. The change in *WWP2* expression (as a surrogate for mir140 expression) was plotted against the average fold change of predicted miRNA targets for each study. Using regression analysis we identified a significant inverse relationship between *WWP2* and miR-140-5p (Figure 6A; r^2=0.38, p=3.08x10^{-14}) and miR-140-3p.1, predicted targets (Figure 6B; r^2=0.16, p=5.46x10^{-6}). There was also minimal but significant inverse relationship between *WWP2* and miR-140-3p.2 predicted targets (Figure 6C; r^2=0.05, p=0.012). Thus, where *WWP2* expression changes, expression of predicted miR-140 targets have a tendency to change in the opposite direction. Next we determined if there was enrichment for miR-140 predicted targets in genes whose expression level changes reciprocally to *WWP2.* Indeed, within studies where *WWP2* expression increased there was generally a larger percentage of predicted miR-140-5p and miR-140-3p.1 targets within down-regulated genes than up-regulated genes (Figure 6E and F). However, when *WWP2* expression decreased, only miR-140-5p predicted targets were enriched in the up-regulated genes (Figure 6E). There was minimal enrichment of miR-140-3p.2 targets (Figure 6G).

To determine if miR-140-5p and miR-140-3p isomiR targets are regulated in a similar way in each study we correlated relative enrichment (enrichment in upregulated divided by enrichment in down-regulated genes) for miR-140-5p, miR-140-3p.1 and miR-140-3p.2 predicted targets. Enrichment observed for miR-140-5p targets correlated with enrichment for miR-140-3p.1 targets (Figure 6I). Studies where miR-140-5p and miR-140-3p.1 predicted targets display a greater enrichment in up-regulated genes than in down-regulated genes (enrichment in up divided by enrichment down >1), were mostly studies where *WWP2* decreased (Figure 6I). Likewise in studies where miR-140-5p and miR-140-3p.1 predicted targets display a greater enrichment in down-regulated genes than up-regulated genes (ratio of up:down-regulated <1), were mostly studies where *WWP2* increased (Figure 6I). These data indicate that when *WWP2* expression increases, both miR-140-5p and miR-140-3p.1 predicted targets are repressed. Furthermore, they also indicate that miR-140-3p.2, essentially the canonical form of miR-140-3p, has a less significant biological effect on gene expression than miR-140-5p or miR-140-3p.1.

To further refine prediction of putative direct miR-140 targets, we investigated the overlap between four criteria; A) TargetScan predicted targets (Figure 2), B) genes down-regulated following overexpression in human chondrocytes (Figure 3), C) genes upregulated in mice lacking the *Mir140* locus (Figure 4) and D) genes up-regulated during MSC chondrogenesis (Figure 5). There were 5, 4 and 0 genes matching all of these criteria for miR-140-5p, miR-140-3p.1 and miR-140-3p.2 respectively (Supplementary Figure 5A, Table S9). *ABCA1* was common to both miR-140-5p and miR-140-3p.1. Furthermore, *ABCA1* was inversely correlated with *WWP2* across multiple skeletal data sets, with the inverse correlation being strongest in cartilage (Supplementary Figure 5B), where miR-140 expression is highest.

DISCUSSION

IsomiRs of miR-140-3p have previously been detected in chondrocytes (36), breast cancer cell lines (37) and endometrial tissue (38), and are now recognised in TargetScan 7.2 (39). In TargetScan and this study, miR-140-3p.2 refers to the original miRBase annotation and miR-140-3p.1 refers to the newly identified isomiR. MiRBase does not currently include specific annotation for isomiRs. In addition to miR-140-3p isomiRs, we detected isomiRs for a number of other miRNAs in cartilage, including miR-455, which is also now recognised in TargetScan 7.2 (39). Sequence analysis of the miR-1246 isomiR, suggests it may in fact be miR-1290. According to miRBase these two miRNAs are transcribed from two separate loci on chromosome 2 and 1, respectively. However, Mazieres *et al.* (40), suggest miR-1246 and miR-1290 are processed from a common transcript (RNU2), which explains their similar sequence and expression pattern during chondrogenesis (9). Many of the other detected cartilage isomiRs can be transcribed from more than one genomic location (hsa-mir-199a-3p, hsa-mir-199b-3p, hsa-mir-29b, hsa-mir-101, hsa-mir-320c, hsa-mir-103b), raising the possibility that each genomic location gives rise to only one mature miRNA, which are then perceived as isomiRs.

Primary miRNA transcripts are processed to pre-miRNAs in the nucleus by DROSHA (a dsRNA nuclear type III endoribonuclease) facilitated by DGCR8, and subsequently cleaved in the cytoplasm by DICER to generate functionally mature miRNAs molecules. IsomiRs are generated because both DICER and DROSHA process miRNA precursors imprecisely generating miRNA variants with several plus/minus nucleotides at the 5' and/or 3' (41). In addition, miRNAs can also be post-transcriptionally adenylated or uridylated, which could modify miRNA targeting properties and/or their stability (42,43). The importance of endogenously expressed 5′ isomiRs has been demonstrated by Chiang *et al*. (30), where they showed deletion of miR-223 in mouse neutrophils (44), resulted in an increase in expression of predicted targets of a lesser expressed 5′ isomiR of the microRNA. Although it is established 5′ isomiRs can regulate gene expression, the functional reason for their presence is still debated. 5′isomiRs of a given microRNA can have highly over-lapping targets, potentially increasing the specificity of regulation of a particular pathways (24,45). However, 5’ isomiRs can also increase a microRNAs target repertoire (46). Manzano *et al.*, suggest these apparent discrepancies are dependent upon the presence or absence of a U at position 2 of the longer isomiR sequence (47). miR-140-3p.2, the longest 5’ isomiR of miR-140-3p, does not possess a U at position 2, which would suggest miR-140-3p.1 and miR-140-3p.2 have divergent target repertoires, decreasing their effect on any single target (47). Although not directly studied here, 3’isomiRs are widespread, with 3’ adenine additions being most common and reported to decrease the targeting ability of the miRNA (42). Some of the miR-140-3p isomiRs with the same seeds as miR-140-3p.1 and miR-140-3p.2 have 3’ adenine additions, which in combination with divergent 5’isomiRs may account for the poor target repression by miR-140-3p.1 and miR-140-3p.2 compared with miR-140-5p.

Interestingly, there is an inverse correlation between the ability to repress targets (miR-140-5p>miR-140-3p.1>miR-140-3p.2) and expression in cartilage (miR-140-3p.2>miR-140-3p.2>miR-140-5p).

A mutation within miR-140-5p has recently been identified as causing a human skeletal dysplasia (48). Mice with the corresponding miR-140-5p mutation phenocopy the human situation and display a loss of miR-140-5p target repression (in rib chondrocytes). This is thought to contribute to the phenotype since the expression of miR-140-3p.1 and miR-140-3.2 were not affected by the mutation. Interestingly though, the phenotypes of the miR-140-5p mutant mice and *MIR140* null mice show some differences, which could perhaps be accounted for by the latter animals lacking miR-140-3p and the isomiRs described here.

Here we identify discrete targets for miR-140-3p.1 and miR-140-3p.2. LIPA, a cholesterol ester hydrolase, is a target of miR-140-3p.1 but not miR-140-3p.2. *LIPA* expression decreases during chondrocyte development in human (49), consistent with high miR-140-3p.1 expression in chondrocytes. SDC4 (Syndecan-4), a target of miR-140-3p.2 but not miR-140-3p.1, has been implicated in osteoarthritis progression through activation of ADAMTS5 (50), suggesting *SDC4* suppression by miR-140-3p.2 may maintain a healthy cartilage phenotype. *HMGCR* and *HMGCS1* have now been identified as targets of miR-140-3p.1 in breast cancer cells (51), however their expression was not repressed by miR-140-3p.1 in human chondrocytes, indicating miR-140 isomiRs act in a tissue-specific manner. Even where repression was determined, the effects of miR-140-3p.1 and miR-140-3p.2 were relatively small (up to ~25% repression) when compared to the effects of miR-140-5p on a previously published miR-140-5p direct target, FZD6 (~60% repression) (9). Furthermore, contra to the expected miR-140-3p.1 repression of targets, miR-140-3p.1 increased luciferase for the *MARCKSL1*, *KIAA1199* and *MAPILC3A* 3'UTR constructs, possibly suggesting stabilisation of transcript, as previously described for miR-322 and *MEK1* (52).

The target repertoire of miR-140-3p.1 and miR-140-3p.2 are vastly different, due to established miRNA target interaction rules. The only shared binding motif is 'CTGTGG', which can be recognised by miR-140-3p.1 as a 6merBS (CTGTGG), or by miR-140-3p.2 as a 7m8-BS or 8mer-BS if followed by T or TA, respectively (CTGTGG**T** or CTGTGG**TA**) (Figure 2A). Indeed the miR-140-3p.1 6merBS (CTGTGG) was the most enriched miR-140-3p.1 motif following miR-140-3p.2 transfection (Figure 3E; red line in bottom right), indicating miR-140-3p isomiRs act according to traditional miRNA targeting rules. Other types of miRNA binding sites such as 3′compensatory binding (53) may be shared between miR-140-3p.1 and miR-140-3p.2, but are less common and more difficult to predict.

Analysis of miRNA target interactions within published skeletal data sets, using the *WWP2* host gene as a surrogate for miRNA expression, along with directly overexpressing the miRNAs and observing changes in gene expression, suggests these types of analyses can

improve miRNA target prediction within skeletal tissues. Caution must however be taken as *WWP2* and *MIR140* can display tissue specific expression (35), furthermore, miRNA are often co-expressed with their targets as part of a larger more complex network (54). This type of co-expression may account for the studies where a correlation, rather than an inverse correlation, between *WWP2* and miR-140 targets was observed. Where an inverse correlation was observed between *WWP2* and miR-140 predicted targets this was greatest within cartilage. Together, these data indicate the miR-140-3p isomiR miR-140-3p.1 can function to down-regulate transcript expression *in vitro* and *in vivo*.

In conclusion, analysis of human cartilage sRNA-seq identified an abundant isomiR to miR-140-3p, miR-140-3p.1. Functional analysis in human and murine chondrocytes and analysis of published skeletal datasets suggests the newly identified isomiR is more effective than the original miR-140-3p annotated in miRBase. These data suggest the function of miR-140-3p, which has previously been largely ignored, should be revisited.

ACCESSION NUMBERS

Microarray data and RNA-seq data are available at NCBI GEO datasets with the accession numbers GSE144374.

SUPPLEMENTARY DATA

Supplementary Data are available online.

ACKNOWLEDGEMENT

pLKO.1-puro U6 sgRNA BfuAI large stuffer was a gift from Scot Wolfe (Addgene plasmid # 52628 ; http://n2t.net/addgene:52628 ; RRID:Addgene_52628).

FUNDING

This work was supported by the Oliver Bird Rheumatism Programme (Nuffield Foundation); Medical Research Council and Versus Arthritis as part of the MRC-Arthritis Research UK Centre for Integrated Research into Musculoskeletal Ageing (CIMA) [JXR 10641, MR/P020941/1]; Versus Arthritis [19424, 22043]; the JGW Patterson Foundation; The Dunhill Medical Trust [R476/0516); and the NIHR Newcastle Biomedical Research.

CONFLICT OF INTEREST

All authors declare no conflicts of interest.

REFERENCES

1. Bartel, D.P. (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*, **116**, 281-297.
2. Griffiths-Jones, S. (2004) The microRNA Registry. *Nucleic acids research*, **32**, D109-111.
3. Bartel, D.P. (2009) MicroRNAs: target recognition and regulatory functions. *Cell*, **136**, 215-233.
4. Lewis, B.P., Burge, C.B. and Bartel, D.P. (2005) Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell*, **120**, 15-20.
5. Morin, R.D., O'Connor, M.D., Griffith, M., Kuchenbauer, F., Delaney, A., Prabhu, A.L., Zhao, Y., McDonald, H., Zeng, T., Hirst, M. *et al.* (2008) Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells. *Genome research*, **18**, 610-621.
6. Wyman, S.K., Knouf, E.C., Parkin, R.K., Fritz, B.R., Lin, D.W., Dennis, L.M., Krouse, M.A., Webster, P.J. and Tewari, M. (2011) Post-transcriptional generation of miRNA variants by multiple nucleotidyl transferases contributes to miRNA transcriptome complexity. *Genome research*, **21**, 1450-1461.
7. Miyaki, S., Nakasa, T., Otsuki, S., Grogan, S.P., Higashiyama, R., Inoue, A., Kato, Y., Sato, T., Lotz, M.K. and Asahara, H. (2009) MicroRNA-140 is expressed in differentiated human articular chondrocytes and modulates interleukin-1 responses. *Arthritis and rheumatism*, **60**, 2723-2730.
8. Miyaki, S., Sato, T., Inoue, A., Otsuki, S., Ito, Y., Yokoyama, S., Kato, Y., Takemoto, F., Nakasa, T., Yamashita, S. *et al.* (2010) MicroRNA-140 plays dual roles in both cartilage development and homeostasis. *Gene Dev*, **24**, 1173-1185.
9. Barter, M.J., Tselepi, M., Gómez, R., Woods, S., Hui, W., Smith, G.R., Shanley, D.P., Clark, I.M. and Young, D.A. (2015) Genome-Wide MicroRNA and Gene Analysis of Mesenchymal Stem Cell Chondrogenesis Identifies an Essential Role and Multiple Targets for miR-140-5p. *Stem cells (Dayton, Ohio)*, **33**, 3266-3280.
10. Crowe, N., Swingler, T.E., Le, L.T., Barter, M.J., Wheeler, G., Pais, H., Donell, S.T., Young, D.A., Dalmay, T. and Clark, I.M. (2016) Detecting new microRNAs in human osteoarthritic chondrocytes identifies miR-3085 as a human, chondrocyte-selective, microRNA. *Osteoarthritis and cartilage / OARS, Osteoarthritis Research Society*, **24**, 534-543.
11. Kozomara, A. and Griffiths-Jones, S. (2011) miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic acids research*, **39**, D152-157.
12. Smyth, G.K. (2004) Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Statistical applications in genetics and molecular biology*, **3**, Article3.
13. Metsalu, T. and Vilo, J. (2015) ClustVis: a web tool for visualizing clustering of multivariate data using Principal Component Analysis and heatmap. *Nucleic acids research*, **43**, W566-570.
14. Agarwal, V., Bell, G.W., Nam, J.W. and Bartel, D.P. (2015) Predicting effective microRNA target sites in mammalian mRNAs. *Elife*, **4**.
15. van Dongen, S., Abreu-Goodger, C. and Enright, A.J. (2008) Detecting microRNA binding and siRNA off-target effects from expression data. *Nature methods*, **5**, 1023-1025.
16. Huang da, W., Sherman, B.T. and Lempicki, R.A. (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*, **4**, 44-57.
17. Ittner, L.M. and Gotz, J. (2007) Pronuclear injection for the production of transgenic mice. *Nat Protoc*, **2**, 1206-1215.
18. Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A. and Zhang, F. (2013) Genome engineering using the CRISPR-Cas9 system. *Nat Protoc*, **8**, 2281-2308.

19. Yang, H., Wang, H., Shivalila, C.S., Cheng, A.W., Shi, L. and Jaenisch, R. (2013) One-step generation of mice carrying reporter and conditional alleles by CRISPR/Cas-mediated genome engineering. *Cell*, **154**, 1370-1379.
20. Grigelioniene, G., Suzuki, H.I., Taylan, F., Mirzamohammadi, F., Borochowitz, Z.U., Ayturk, U.M., Tzur, S., Horemuzova, E., Lindstrand, A., Weis, M.A. *et al.* (2019) Gain-of-function mutation of microRNA-140 in human skeletal dysplasia. *Nat Med*.
21. Patro, R., Duggal, G., Love, M.I., Irizarry, R.A. and Kingsford, C. (2017) Salmon provides fast and bias-aware quantification of transcript expression. *Nature methods*, **14**, 417-419.
22. Risso, D., Ngai, J., Speed, T.P. and Dudoit, S. (2014) Normalization of RNA-seq data using factor analysis of control genes or samples. *Nat Biotechnol*, **32**, 896-902.
23. Love, M.I., Huber, W. and Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology*, **15**, 550.
24. Cloonan, N., Wani, S., Xu, Q., Gu, J., Lea, K., Heater, S., Barbacioru, C., Steptoe, A.L., Martin, H.C., Nourbakhsh, E. *et al.* (2011) MicroRNAs and their isomiRs function cooperatively to target common biological pathways. *Genome biology*, **12**, R126.
25. Stark, M.S., Tyagi, S., Nancarrow, D.J., Boyle, G.M., Cook, A.L., Whiteman, D.C., Parsons, P.G., Schmidt, C., Sturm, R.A. and Hayward, N.K. (2010) Characterization of the Melanoma miRNAome by Deep Sequencing. *PLoS ONE*, **5**, e9685.
26. Witten, D., Tibshirani, R., Gu, S.G., Fire, A. and Lui, W.O. (2010) Ultra-high throughput sequencing-based small RNA discovery and discrete statistical biomarker analysis in a collection of cervical tumours and matched controls. *Bmc Biol*, **8**.
27. Kuchen, S., Resch, W., Yamane, A., Kuo, N., Li, Z., Chakraborty, T., Wei, L., Laurence, A., Yasuda, T., Peng, S. *et al.* (2010) Regulation of microRNA expression and abundance during lymphopoiesis. *Immunity*, **32**, 828-839.
28. Flores, O., Kennedy, E.M., Skalsky, R.L. and Cullen, B.R. (2014) Differential RISC association of endogenous human microRNAs predicts their inhibitory potential. *Nucleic acids research*, **42**, 4629-4639.
29. Burroughs, A.M., Ando, Y., de Hoon, M.J., Tomaru, Y., Suzuki, H., Hayashizaki, Y. and Daub, C.O. (2011) Deep-sequencing of human Argonaute-associated small RNAs provides insight into miRNA sorting and reveals Argonaute association with RNA fragments of diverse origin. *RNA Biol*, **8**, 158-177.
30. Chiang, H.R., Schoenfeld, L.W., Ruby, J.G., Auyeung, V.C., Spies, N., Baek, D., Johnston, W.K., Russ, C., Luo, S., Babiarz, J.E. *et al.* (2010) Mammalian microRNAs: experimental evaluation of novel and previously annotated genes. *Gene Dev*, **24**, 992-1009.
31. Soul, J., Hardingham, T., Boot-Handford, R. and Schwartz, J.M. (2018) SkeletalVis: An exploration and meta-analysis data portal of cross-species skeletal transcriptomics data. *Bioinformatics*.
32. Franca, G.S., Vibranovski, M.D. and Galante, P.A. (2016) Host gene constraints and genomic context impact the expression and evolution of human microRNAs. *Nat Commun*, **7**, 11438.
33. Gennarino, V.A., Sardiello, M., Avellino, R., Meola, N., Maselli, V., Anand, S., Cutillo, L., Ballabio, A. and Banfi, S. (2009) MicroRNA target prediction by expression analysis of host genes. *Genome research*, **19**, 481-490.
34. Radfar, M.H., Wong, W. and Morris, Q. (2011) Computational prediction of intronic microRNA targets using host gene expression reveals novel regulatory mechanisms. *PLoS One*, **6**, e19312.
35. Yamashita, S., Miyaki, S., Kato, Y., Yokoyama, S., Sato, T., Barrionuevo, F., Akiyama, H., Scherer, G., Takada, S. and Asahara, H. (2012) L-Sox5 and Sox6 proteins enhance chondrogenic miR-140 microRNA expression by strengthening dimeric Sox9 activity. *The Journal of biological chemistry*, **287**, 22206-22215.

36. Haseeb, A., Makki, M.S., Khan, N.M., Ahmad, I. and Haqqi, T.M. (2017) Deep sequencing and analyses of miRNAs, isomiRs and miRNA induced silencing complex (miRISC)-associated miRNome in primary human chondrocytes. *Scientific reports*, **7**, 15178.
37. Salem, O., Erdem, N., Jung, J., Munstermann, E., Worner, A., Wilhelm, H., Wiemann, S. and Korner, C. (2016) The highly expressed 5'isomiR of hsa-miR-140-3p contributes to the tumor-suppressive effects of miR-140 by reducing breast cancer proliferation and migration. *BMC Genomics*, **17**, 566.
38. Krawczynski, K., Bauersachs, S., Reliszko, Z.P., Graf, A. and Kaczmarek, M.M. (2015) Expression of microRNAs and isomiRs in the porcine endometrium: implications for gene regulation at the maternal-conceptus interface. *BMC Genomics*, **16**, 906.
39. Agarwal, V., Bell, G.W., Nam, J.W. and Bartel, D.P. (2015) Predicting effective microRNA target sites in mammalian mRNAs. *elife*, **4**, e05005.
40. Mazieres, J., Catherinne, C., Delfour, O., Gouin, S., Rouquette, I., Delisle, M.B., Prevot, G., Escamilla, R., Didier, A., Persing, D.H. *et al.* (2013) Alternative processing of the U2 small nuclear RNA produces a 19-22nt fragment with relevance for the detection of non-small cell lung cancer in human serum. *PLoS ONE*, **8**, e60134.
41. Neilsen, C.T., Goodall, G.J. and Bracken, C.P. (2012) IsomiRs--the overlooked repertoire in the dynamic microRNAome. *Trends Genet*, **28**, 544-549.
42. Burroughs, A.M., Ando, Y., de Hoon, M.J., Tomaru, Y., Nishibu, T., Ukekawa, R., Funakoshi, T., Kurokawa, T., Suzuki, H., Hayashizaki, Y. *et al.* (2010) A comprehensive survey of 3' animal miRNA modification events and a possible role for 3' adenylation in modulating miRNA targeting effectiveness. *Genome research*, **20**, 1398-1410.
43. Marzi, M.J., Ghini, F., Cerruti, B., de Pretis, S., Bonetti, P., Giacomelli, C., Gorski, M.M., Kress, T., Pelizzola, M., Muller, H. *et al.* (2016) Degradation dynamics of microRNAs revealed by a novel pulse-chase approach. *Genome research*, **26**, 554-565.
44. Baek, D., Villen, J., Shin, C., Camargo, F.D., Gygi, S.P. and Bartel, D.P. (2008) The impact of microRNAs on protein output. *Nature*, **455**, 64-71.
45. Llorens, F., Bañez-Coronel, M. and Pantano, L. (2013) A highly expressed miR-101 isomiR is a functional silencing small RNA. *BMC …*.
46. Tan, G., Chan, E., Molnar, A., Sarkar, R., Alexieva, D., Isa, I., Robinson, S., Zhang, S., Ellis, P. and Langford, C.F. (2014) 5 isomiR variation is of functional and evolutionary importance. *Nucleic acids research*, **42**, 9424-9435.
47. Manzano, M., Forte, E., Raja, A.N., Schipma, M.J. and Gottwein, E. (2015) Divergent target recognition by coexpressed 5'-isomiRs of miR-142-3p and selective viral mimicry. *Rna*, **21**, 1606-1620.
48. Grigelioniene, G., Suzuki, H.I., Taylan, F., Mirzamohammadi, F., Borochowitz, Z.U., Ayturk, U.M., Tzur, S., Horemuzova, E., Lindstrand, A., Weis, M.A. *et al.* (2019) Gain-of-function mutation of microRNA-140 in human skeletal dysplasia. *Nat Med*, **25**, 583-590.
49. Wu, L., Bluguermann, C., Kyupelyan, L., Latour, B., Gonzalez, S., Shah, S., Galic, Z., Ge, S.D., Zhu, Y.H., Petrigliano, F.A. *et al.* (2013) Human Developmental Chondrogenesis as a Basis for Engineering Chondrocytes from Pluripotent Stem Cells. *Stem Cell Rep*, **1**, 575-589.
50. Echtermeyer, F., Bertrand, J., Dreier, R., Meinecke, I., Neugebauer, K., Fuerst, M., Lee, Y.J., Song, Y.W., Herzog, C., Theilmeier, G. *et al.* (2009) Syndecan-4 regulates ADAMTS-5 activation and cartilage breakdown in osteoarthritis. *Nat Med*, **15**, 1072-1076.
51. Bhardwaj, A. and Cancer …, S.-H. (2018) The isomiR-140-3p-regulated mevalonic acid pathway as a potential target for prevention of triple negative breast cancer. *Breast Cancer …*.
52. Bluhm, B., Ehlen, H.W.A., Holzer, T., Georgieva, V.S., Heilig, J., Pitzler, L., Etich, J., Bortecen, T., Frie, C., Probst, K. *et al.* (2017) miR-322 stabilizes MEK1 expression to inhibit RAF/MEK/ERK pathway activation in cartilage. *Development*, **144**, 3562-3577.

53. Grimson, A., Farh, K.K., Johnston, W.K., Garrett-Engele, P., Lim, L.P. and Bartel, D.P. (2007) MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Molecular cell*, **27**, 91-105.
54. Cora, D., Re, A., Caselle, M. and Bussolino, F. (2017) MicroRNA-mediated regulatory circuits: outlook and perspectives. *Phys Biol*, **14**.

MAIN FIGURE LEGENDS

Figure 1. miR-140-3p isomiRs are abundantly expressed in cartilage. (A) Pie chart illustrating the relative abundance of all miRNAs expressed in cartilage, data combines isomiR and canonical reads for each miRNA. (B) Pie chart illustrating the isomiR type for all reads aligned to miR-140-3p. (C) Pie chart illustrating seed sequences for all reads aligned to miR-140-3p, >99% of reads have one of two seed sequences; ACCACAG (blue) and CCACAGG (red). (D) Pie chart of the percentage of templated and non-templated reads for miR-140-3p within cartilage. (E) Histogram of all sequencing reads aligned to miR-140-3p in cartilage. (F) Histogram of sequencing reads aligned to miR-140-3p, whose 5' end is as indicated in miRBase. (G) Histogram of sequencing reads aligned to miR-140-3p, whose 5' end is one nucleotide shorter than indicated in miRBase. (H) Bar chart indicting the sequences and number of reads that contribute to each of the two seed sequences. Individual bars represent total read number from 3 separate individuals. (I) qRT-PCR for miR-140-3p.1 and miR-140-3p.2. Assays designed to detect each isomiR are able to distinguish between miR-140-3p.1 and miR-140-3p.2 spike-ins (2 fold serial dilution) and detect high expression of each isomiR in cartilage (2 fold serial dilution).

Figure 2. miR-140-3p.1 and miR-140-3p.2 are predicted to have distinct targets. (A) Schematic illustrating the sequences of miRBase miR-140-3p, miR-140-3p.1, miR-140-3p.2 and miR-140-5p and their potential binding sites within target mRNAs using the TargetScan categorisation. (B) Venn diagram illustrating little cross over of targets predicted by TargetScan 7.2 for miR-140-3p.1, miR-140-3p.2 and miR-140-5p.

Figure 3. Identification and validation of miR-140-3p.1 and miR-140-3p.2 targets in human chondrocytes. (A) qRT-PCR for miR-140-3p.1 and miR-140-3p.2 following transfection of either control mimic, miR-140-3p.1 mimic or miR-140-3p.2 mimic for 48h. (B) PCA of whole genome micro-array following transfection of mimics described in (A), (140.1= miR-140-3p.1 and 140.2=miR-140-3p.2). (C) Heat map of whole genome micro-array following transfection of mimics as described in (A) (140.1= miR-140-3p.1 and 140.2=miR-140-3p.2). (D) Venn diagram indicating the number of genes down-regulated following overexpression of miR-140-3p.1 and miR-140-3p.2. (E) Enrichment analysis using Sylamer for 3′UTR mRNA motifs complementary to either miR-140-3p.1 or miR-140-3p.2 seed sequences in genes whose expression decreased following transfection of either miR-140-3p mimic, miR-140-3p.1 mimic or miR-140-3p.2 mimic. (F) Gene expression changes of miR-140-3p.1 (only.1;red),

miR-140-3p.2 (only.2;blue) or miR-140-3p.1 and miR-140-3p.2 (common;green) predicted targets following transfection of either miR-140-3p.1 mimic or miR-140-3p.2 mimic. (G-I) 3′ UTR luciferase reporters for miR-140-3p.1 and miR-140-3p.2 putative targets. *, ** or *** indicate level of significance (p<0.05, 0.01 or 0.001, respectively) of either miR-140-3p.1 or miR-140-3p.2 to control. # indicates level of significance (p<0.05, 0.01 or 0.001, respectively) between miR-140-3p.1 and miR-140-3p.2.

Figure 4. Evidence for functional miR-140-3p.1 and miR-140-3p.2 in mouse cartilage. (A) Pie chart illustrating the fraction of miR-140-3p reads with canonical 5′ and 3′ ends (grey), canonical 5′ and either shorter or longer 3′ (blue), canonical 3′ and either shorter or longer 5′ (green), either shorter or longer 5′ and either shorter or longer 3′ (red). (B) As in (A) with miR-140-3p.1 and miR-140-3p.2 indicated in dark red and dark blue respectively. (C) Bar chart indicting number of reads that contribute to (B). (D) % of genes that are predicted targets of either a miR-140-5p (green), miR-140-3p.1 (red) or miR-140-3p.2 (blue), in genes whose expression decrease, does not change or increase following KO of miR-140 in mice. (E) Median fold change (FC) of all genes that are predicted targets of either a miR-140-5p (green), miR-140-3p.1 (red) or miR-140-3p.2 (blue) following KO of miR-140 in mice. (F) Cumulative fraction plot of all genes (black), miR-140-5p (green), miR-140-3p.1 (red) or miR-140-3p.2 (blue) predicted targets, for genes ordered from most decreased expression to most increased expression in Mir140 null mice. Inset: Fraction at a cut-off of log2FC -1. (G) Sylamer analysis for Mir140 null mice; the 100 most enriched 7mers are shown with 7m8 binding sites for miR-140-5p (green), miR-140-3p.1 (red) and miR-140-3p.2 (blue) highlighted.

Figure 5. Evidence for miR-140-3p.1 and miR-140-3p.2 function during human MSC chondrogenesis. Re-analysis of data from Barter *et al* [10] for human MSC differentiation to cartilage. (A) Changes in gene expression for eight MSC and chondrogenic markers. (B) Mean fold change of all genes that are predicted targets of either a miR-140-5p (green), miR-140-3p.1 (red) or miR-140-3p.2 (blue) during MSC chondrogenesis. (C) Sylamer analysis for a miR-140-5p (green), miR-140-3p.1 (red) and miR-140-3p.2 (blue) 7m8 seed sequences.

Figure 6. Correlation of mir140 host gene WWP2 with mir140 target genes in SkeletalVis transcriptomic datasets. (A-D) Correlation analysis of WWP2 and predicted miRNA target expression in 124 pre-existing transcriptome-wide comparisons (both human and mouse). Scatter graph of WWP2 log2FC plotted against the average log2FC for genes predicted to be either a target of miR-140-5p (A), miR-140-3p.1 (B), miR-140-3p.2 (C) or 500 random genes (D), p-value calculated from linear regression analysis. (E-H) Enrichment analysis of miR-140-5p (E), miR-140-3p.1 (F) or miR-140-3p.2 (G) predicted targets and 500 random genes (H) in data sets where WWP2 expression significantly changes. Data is plotted as percentage of targets within up-regulated, no change or down-regulated transcripts against log2FC WWP2. Large icons represent enrichment of miR-140-5p, miR-140-3p.1 and miR-140-3p.2 predicted

targets in up- and down-regulated genes within rib cartilage of MIR140 null mice, where Wwp2 expression does not change. (I-K) Correlation of miR-140-5p and miR-140-3p.1 target enrichment. In studies where WWP2 significantly decreased (blue) and increased (red). The enrichment of miR-140-5p and miR-140-3p.1 and miR-140-3p.2 targets in significantly down and up-regulated genes was calculated. From this the relative enrichment in up vs down-regulated genes was calculated (enrichment in up/ enrichment in down). The relative enrichment for each miRNAs targets were plotted against each other.

SUPPLEMENTARY FIGURE LEGENDS

Supplementary figure 1. Cartilage miRNAs which have abundant 5' isomiRs. Twenty-nine miRNAs with isomiRs that account for more than 5% of the reads for that miRNA and the number of isomiR reads is greater than 100. miR-1246 isomiR (read count 99) is also shown. 5' start position of each isomiR relative to the miRBase annotation is shown: the addition of a nucleotide is designated as '+1' and the loss as '-1', '0' represents reads of the miRBase annotation. Histogram indicates read count.

Supplementary figure 2. Analysis of published sRNA-seq. miR-140-3p.1 and miR-140-3p.2 were present in multiple sRNA-seq data including; melanoma (Stark *et al*. 2010 (25)), cervix (Witten *et al*. 2010 (26)), lymphocytes (Kuchen *et al*. 2010 (27),) and following immunoprecipitation of Argonaute (AGO) proteins (Burroughs *et al*. (29)). Sequences were designated either miR-140-3p.1 or miR-140-3p.2 based upon seed sequence.

Supplementary figure 3 Validation of miR-140 null mouse model and rib chondrocyte RNA-seq. (A) Deletion of Mir-140 was attained using the CRISPR/-Cas9 system. The miR-140 locus was targeted in mouse zygotes using two crRNAs (blue line, protospacer adjacent motif (PAM) shown in red). This successfully deleted the 140 locus sequence depicted (KO mir140). (B) Mice were bred to homozygosity with the genotype in wild-type, heterozygous and null mice confirmed through PCR of ear-notch genomic RNA. (C) Total RNA was extracted from rib chondrocytes of postnatal day 7 (P7) mice, reverse transcribed to cDNA and subjected to real time qRT-PCR analysis for gene expression of miR-140-5p and miR-140-3p (canonical). Values were normalised to U6 and plotted as mean ± standard error of the mean (SEM) (number of samples, WT=7, miR-140$^{-/-}$ =6). *** $p \leq 0.001$. (D) The RNA from several of the same samples as (C) was subjected to RNA-seq (n=4 for WT and miR-140$^{-/-}$). Principle component analysis of length-scaled TPM (Transcripts Per Million) values segregated WT and KO (blue and red). (E) Heat map of the 1200 most differentially expressed genes.

Supplementary figure 4 IsomiR read count and enrichment score for all possible sequences derived from miR-140. (A) Read count for all 5'isomiRs encoded from the mouse miR140 stem loop according to miRBase. (B) Sylamer analysis of rib cartilage RNA from mice lacking miR140 stem loop for 7m8 seed binding sites corresponding to every possible miRNA produced from the miR140 locus. Histogram represents average enrichment score for each seed binding motif.

Supplementary figure 5. Overlap between human and mouse mir140 datasets. (A) overlap between predicted targets, genes down-regulated following overexpression in human articular chondrocytes (HAC), genes up-regulated in mice lacking the miR140 locus and

genes downregulated during human MSC chondrogenesis. (B) Regression analysis of ABCA1 expression and WWP in multiple skeletal tissues, separated based on tissue type.

SUPPLEMENTARY TABLE LEGENDS

Supplementary table 1. Cartilage sRNA-seq. Cartilage sRNA-seq from 3 donors. All sequences including isomiRs aligning to 990 mature miRNAs are shown with read count.

Supplementary table 2. Predicted targets. Predicted targets of miR-140-5p, miR-140-3p.1 and miR-140-3p.2 by TargetScanHuman 7.2. Gene lists correspond to Venn diagram in figure 2B.

Supplementary table 3. Human chondrocyte gene expression. Human chondrocyte gene expression changes following transfection of miRNA mimics: miR-140-3p.1 vs CON, miR-140-3p.2 vs CON, miR-140-3p.1 vs miR-140-3p.2

Supplementary table 4. Human chondrocyte gene list. Gene lists from human chondrocytes: Genes decreased with miR-140-3p.1 vs control, Genes decreased with miR-140-3p.2 vs control, Genes decreased with both miR-140-3p.1 and miR-140-3p.2 vs control, and genes significantly different between miR-140-3p.1 and miR-140-3p.2 when directly compared.

Supplementary table 5. Human chondrocyte pathway analysis. Pathway analysis using DAVID for genes decreased with miR-140-3p.1, genes decreased with miR-140-3p.2 genes significantly different between miR-140-3p.1 and miR-140-3p.2.

Supplementary table 6. MIR140 KO mice gene expression. RNA-seq of mouse rib cartilage of MIR140 KO mice and WT control mice

Supplementary table 7. MIR140 KO mice gene lists. List of up and down-regulated genes in MIR140 KO mice

Supplementary table 8. MIR140 KO mice pathway analysis. Pathway analysis using DAVID for genes up-regulated in MIR140 KO mice

Supplementary table 9. Human vs mouse. Cross over between predicted targets, human chondrocyte data and mouse KO data

Supplementary table 10. Oligonucleotides. Table of oligonucleotides used in this study. Targets of miR-140-3p.1 and miR-140-3p.2 used for cloning into pMiR-GLO by InFusion. Primers used for the generation and genotyping of miR-140$^{-/-}$ mice.

Figure 1

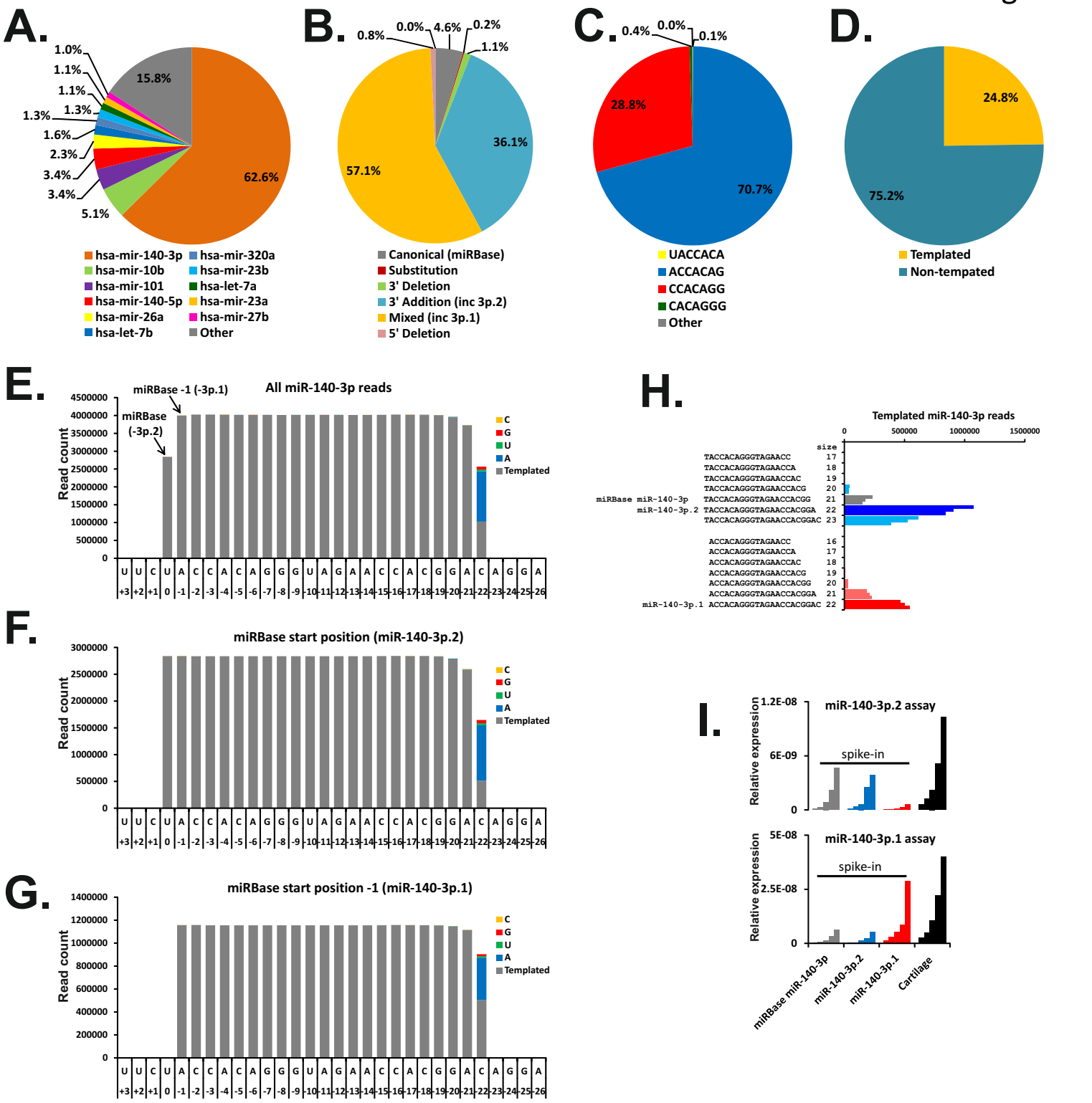

A.
62.6%
15.8%
5.1%
3.4%
3.4%
2.3%
1.6%
1.3%
1.3%
1.1%
1.1%
1.0%
hsa-mir-140-3p
hsa-mir-320a
hsa-mir-10b
hsa-mir-23b
hsa-mir-101
hsa-let-7a
hsa-mir-140-5p
hsa-mir-23a
hsa-mir-26a
hsa-mir-27b
hsa-let-7b
Other
B.
57.1%
36.1%
1.1%
0.2%
4.6%
0.0%
0.8%
Canonical (miRBase)
Substitution
3' Deletion
3' Addition (inc 3p.2)
Mixed (inc 3p.1)
5' Deletion
C.
70.7%
28.8%
0.4%
0.0%
0.1%
UACCACA
ACCACAG
CCACAGG
CACAGGG
Other
D.
75.2%
24.8%
Templated
Non-tempated
E.
All miR-140-3p reads
miRBase -1 (-3p.1)
miRBase (-3p.2)
Read count
C
G
U
A
Templated
F.
miRBase start position (miR-140-3p.2)
Read count
G.
miRBase start position -1 (miR-140-3p.1)
Read count
H.
Templated miR-140-3p reads
size
miRBase miR-140-3p
miR-140-3p.2
miR-140-3p.1
I.
miR-140-3p.2 assay
miR-140-3p.1 assay
spike-in
Relative expression
miRBase miR-140-3p
miR-140-3p.2
miR-140-3p.1
Cartilage

Figure 2

A.

```
        miRNA position:12345678                                  miRNA position:87654321
miRBase miR-140-3p 5' UACCACAGGGUAGAACCACGG 3'        miR-140-3p   8merBS:      CUGUGGUA
                3'nnnnnUGGUGUCnnnnnnnnnnnnnnnnnnn 5'  miR-140-3p   7mer-m8BS:   CUGUGGU
                                                      miR-140-3p   7mer-1aBS:    UGUGGUA
                                                      miR-140-3p   6merBS:       UGUGGU

        miRNA position:12345678                                  miRNA position:87654321
miR-140-3p.1 5' _ACCACAGGGUAGAACCACGGAC 3'            miR-140-3p.1 8merBS:      CCUGUGGA
                3'nnnnnnGGUGUCCnnnnnnnnnnnnnnnnnnn 5'  miR-140-3p.1 7mer-m8BS:   CCUGUGG
                                                      miR-140-3p.1 7mer-1aBS:    CUGUGGA
                                                      miR-140-3p.1 6merBS:       CUGUGG

        miRNA position:12345678                                  miRNA position:87654321
miR-140-3p.2  5'UACCACAGGGUAGAACCACGGA 3'             miR-140-3p.2 8merBS:      CUGUGGUA
                3'nnnnnUGGUGUCnnnnnnnnnnnnnnnnnnn 5'  miR-140-3p.2 7mer-m8BS:   CUGUGGU
                                                      miR-140-3p.2 7mer-1aBS:    UGUGGUA
                                                      miR-140-3p.2 6merBS:       UGUGGU

        miRNA position:12345678                                  miRNA position:87654321
miR-140-5p  5'  CAGUGGUUUUACCCUAUGGUAG 3'             miR-140-5p   8merBS:      AACCACUA
                3'nnnnnUCACCAAnnnnnnnnnnnnnnnnnnn 5'  miR-140-5p   7mer-m8BS:   AACCACU
                                                      miR-140-5p   7mer-1aBS:    ACCACUA
                                                      miR-140-5p   6merBS:       ACCACU
```

B.

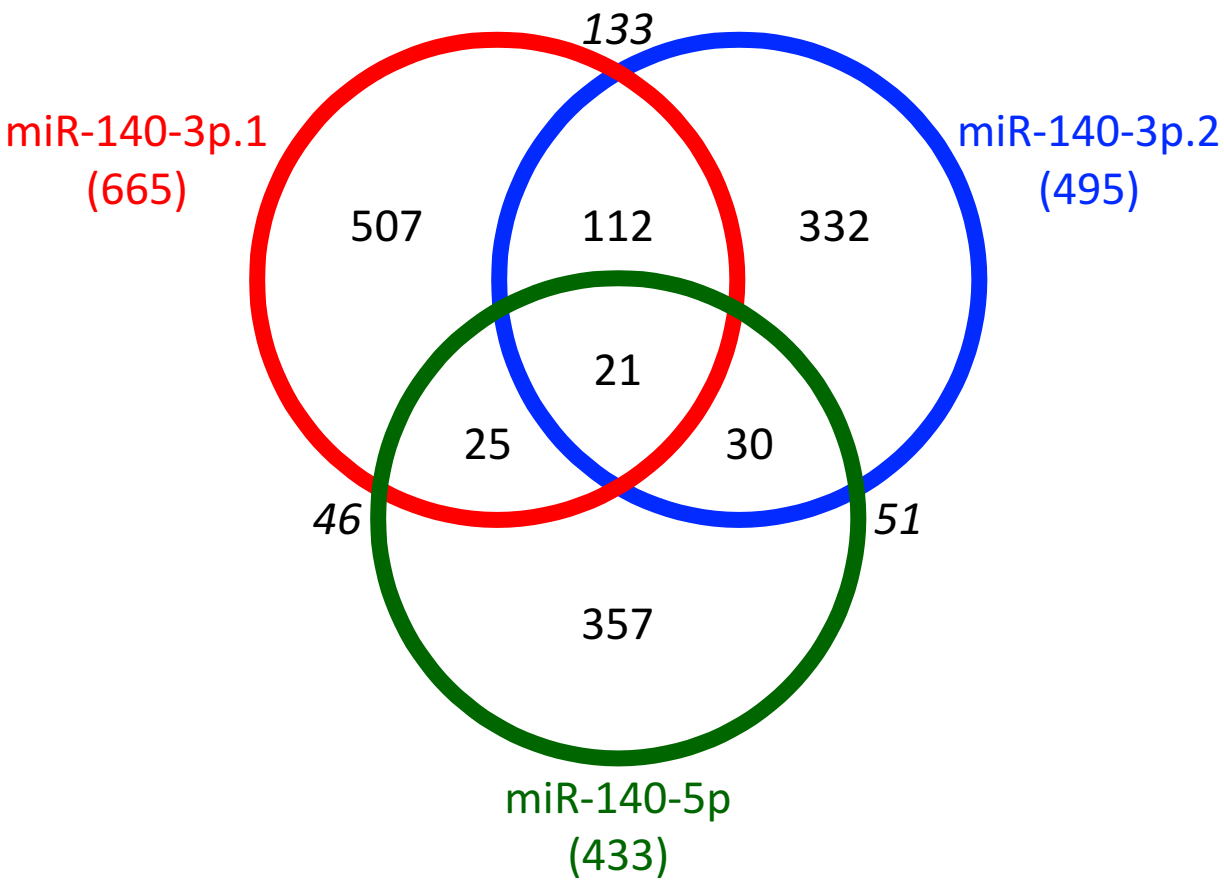


Figure 3

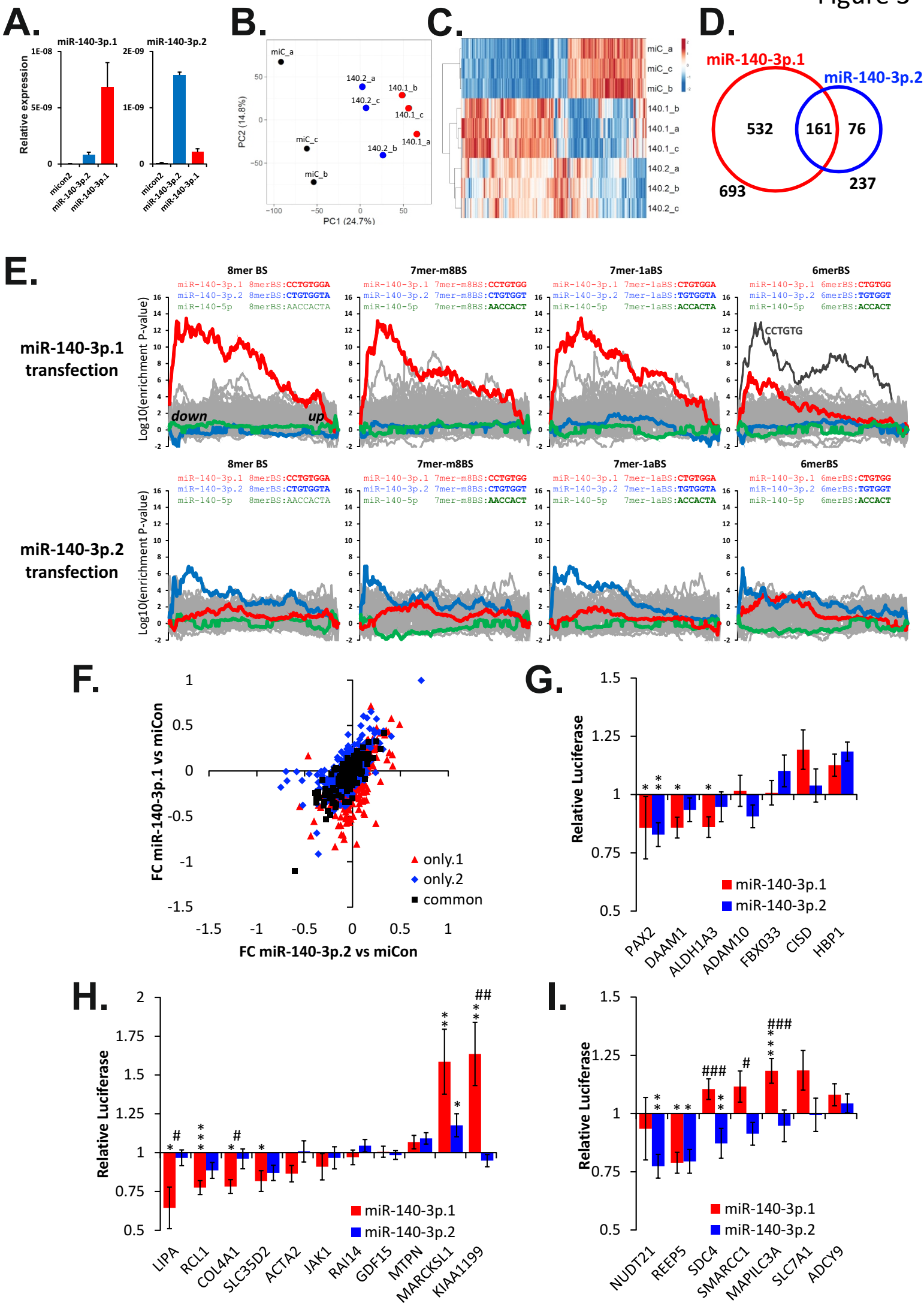

A.
miR-140-3p.1
miR-140-3p.2
Relative expression
1E-08
5E-09
0
2E-09
1E-09
micon2
miR-140-3p.2
miR-140-3p.1
B.
PC2 (14.8%)
PC1 (24.7%)
miC_a
miC_c
miC_b
140.2_a
140.1_b
140.2_c
140.1_c
140.1_a
140.2_b
C.
D.
miR-140-3p.1
miR-140-3p.2
532
161
76
693
237
E.
8mer BS
7mer-m8BS
7mer-1aBS
6merBS
miR-140-3p.1 8merBS:CCTGTGGA
miR-140-3p.2 8merBS:CTGTGGTA
miR-140-5p 8merBS:AACCACTA
miR-140-3p.1 7mer-m8BS:CCTGTGG
miR-140-3p.2 7mer-m8BS:CTGTGGT
miR-140-5p 7mer-m8BS:AACCACT
miR-140-3p.1 7mer-1aBS:CTGTGGA
miR-140-3p.2 7mer-1aBS:TGTGGTA
miR-140-5p 7mer-1aBS:ACCACTA
miR-140-3p.1 6merBS:CTGTGG
miR-140-3p.2 6merBS:TGTGGT
miR-140-5p 6merBS:ACCACT
miR-140-3p.1 transfection
miR-140-3p.2 transfection
Log10(enrichment P-value)
down
up
CCTGTG
F.
FC miR-140-3p.1 vs miCon
FC miR-140-3p.2 vs miCon
only.1
only.2
common
G.
Relative Luciferase
miR-140-3p.1
miR-140-3p.2
PAX2
DAAM1
ALDH1A3
ADAM10
FBXO33
CISD
HBP1
H.
Relative Luciferase
miR-140-3p.1
miR-140-3p.2
LIPA
RCL1
COL4A1
SLC35D2
ACTA2
JAK1
RAI14
GDF15
MTPN
MARCKSL1
KIAA1199
I.
Relative Luciferase
miR-140-3p.1
miR-140-3p.2
NUDT21
REEP5
SDC4
SMARCC1
MAPILC3A
SLC7A1
ADCY9

Figure 4

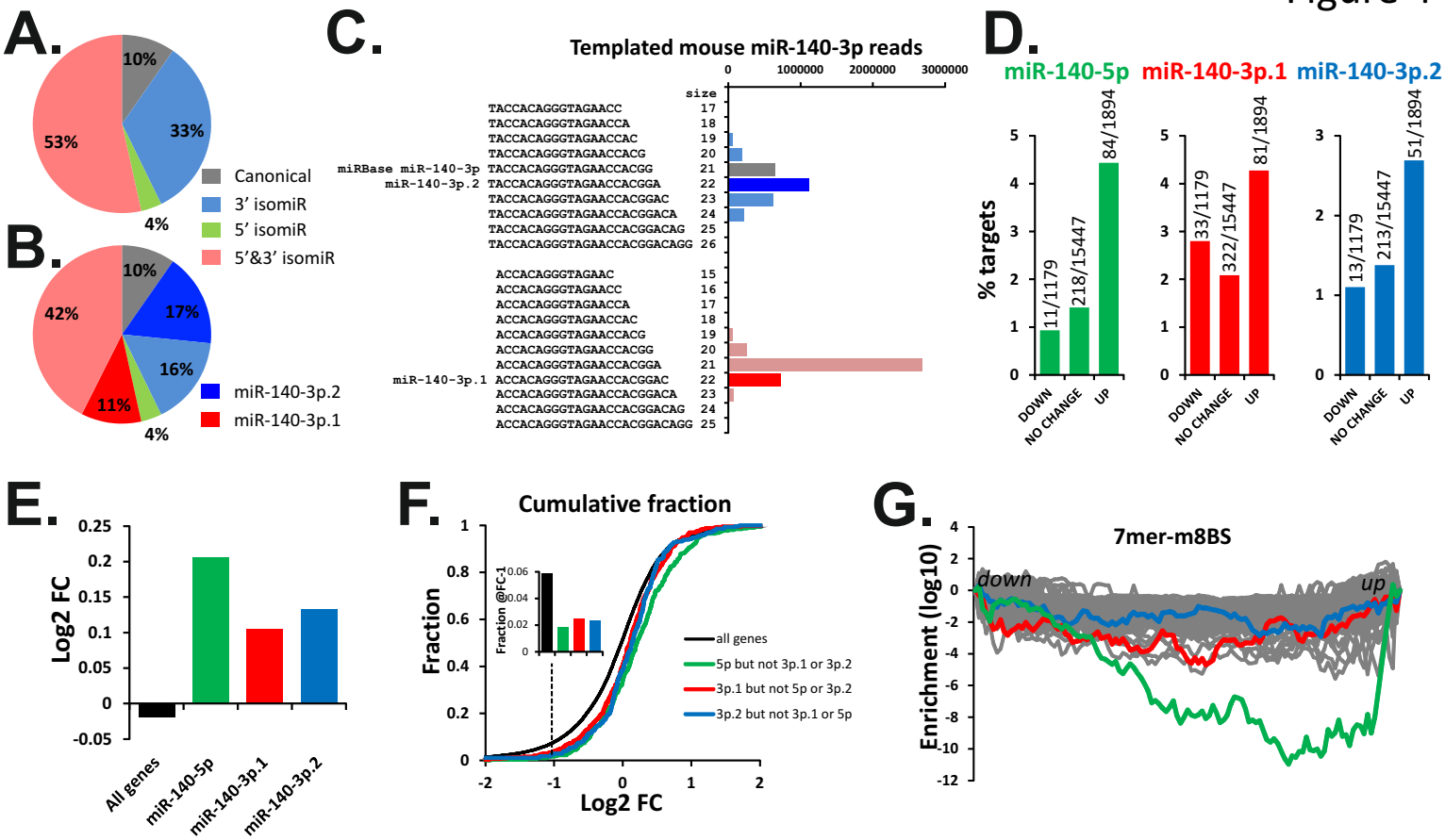

A.
10%
33%
53%
4%
Canonical
3' isomiR
5' isomiR
5'&3' isomiR
B.
10%
17%
42%
16%
11%
4%
miR-140-3p.2
miR-140-3p.1
C.
Templated mouse miR-140-3p reads
0
1000000
2000000
3000000
size
TACCACAGGGTAGAACC 17
TACCACAGGGTAGAACCA 18
TACCACAGGGTAGAACCAC 19
TACCACAGGGTAGAACCACG 20
miRBase miR-140-3p TACCACAGGGTAGAACCACGG 21
miR-140-3p.2 TACCACAGGGTAGAACCACGGA 22
TACCACAGGGTAGAACCACGGAC 23
TACCACAGGGTAGAACCACGGACA 24
TACCACAGGGTAGAACCACGGACAG 25
TACCACAGGGTAGAACCACGGACAGG 26
ACCACAGGGTAGAAC 15
ACCACAGGGTAGAACC 16
ACCACAGGGTAGAACCA 17
ACCACAGGGTAGAACCAC 18
ACCACAGGGTAGAACCACG 19
ACCACAGGGTAGAACCACGG 20
ACCACAGGGTAGAACCACGGA 21
miR-140-3p.1 ACCACAGGGTAGAACCACGGAC 22
ACCACAGGGTAGAACCACGGACA 23
ACCACAGGGTAGAACCACGGACAG 24
ACCACAGGGTAGAACCACGGACAGG 25
D.
miR-140-5p
miR-140-3p.1
miR-140-3p.2
% targets
11/1179
218/15447
84/1894
33/1179
322/15447
81/1894
13/1179
213/15447
51/1894
DOWN
NO CHANGE
UP
E.
Log2 FC
All genes
miR-140-5p
miR-140-3p.1
miR-140-3p.2
F.
Cumulative fraction
Fraction
Fraction @FC-1
Log2 FC
all genes
5p but not 3p.1 or 3p.2
3p.1 but not 5p or 3p.2
3p.2 but not 3p.1 or 5p
G.
7mer-m8BS
Enrichment (log10)
down
up

Figure 5

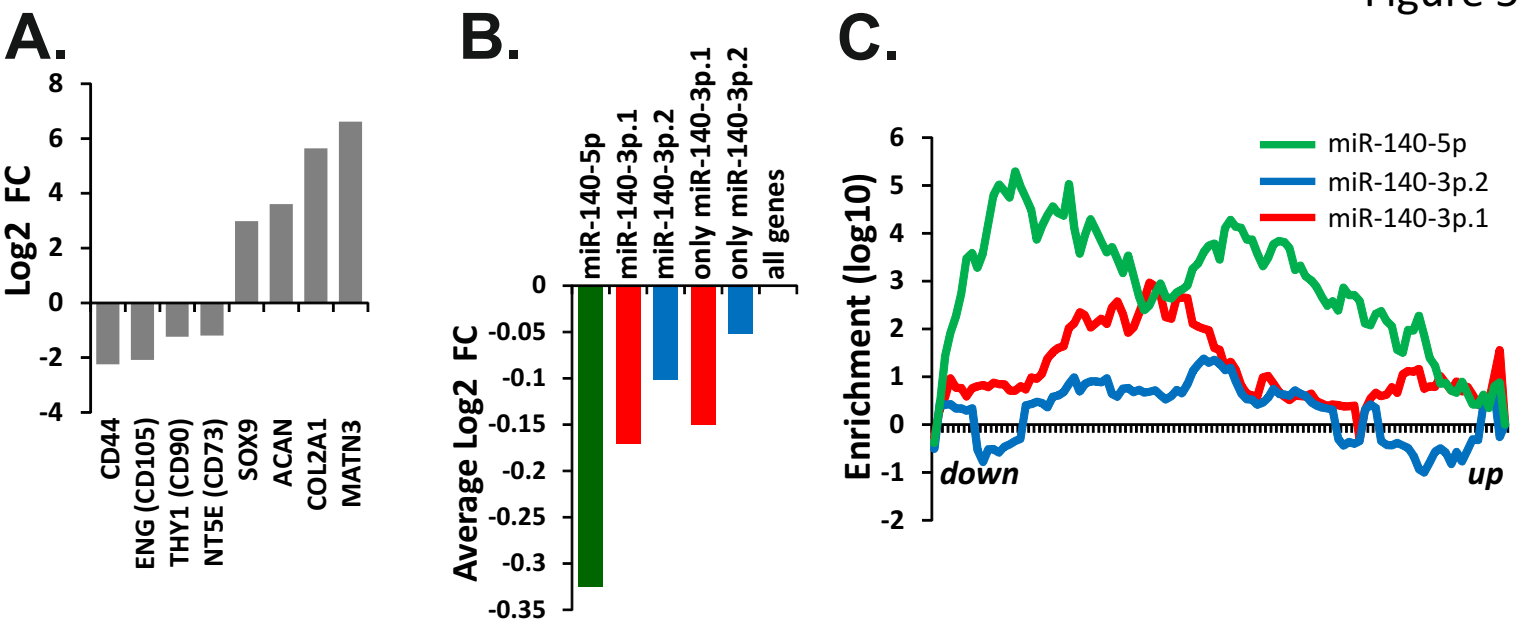
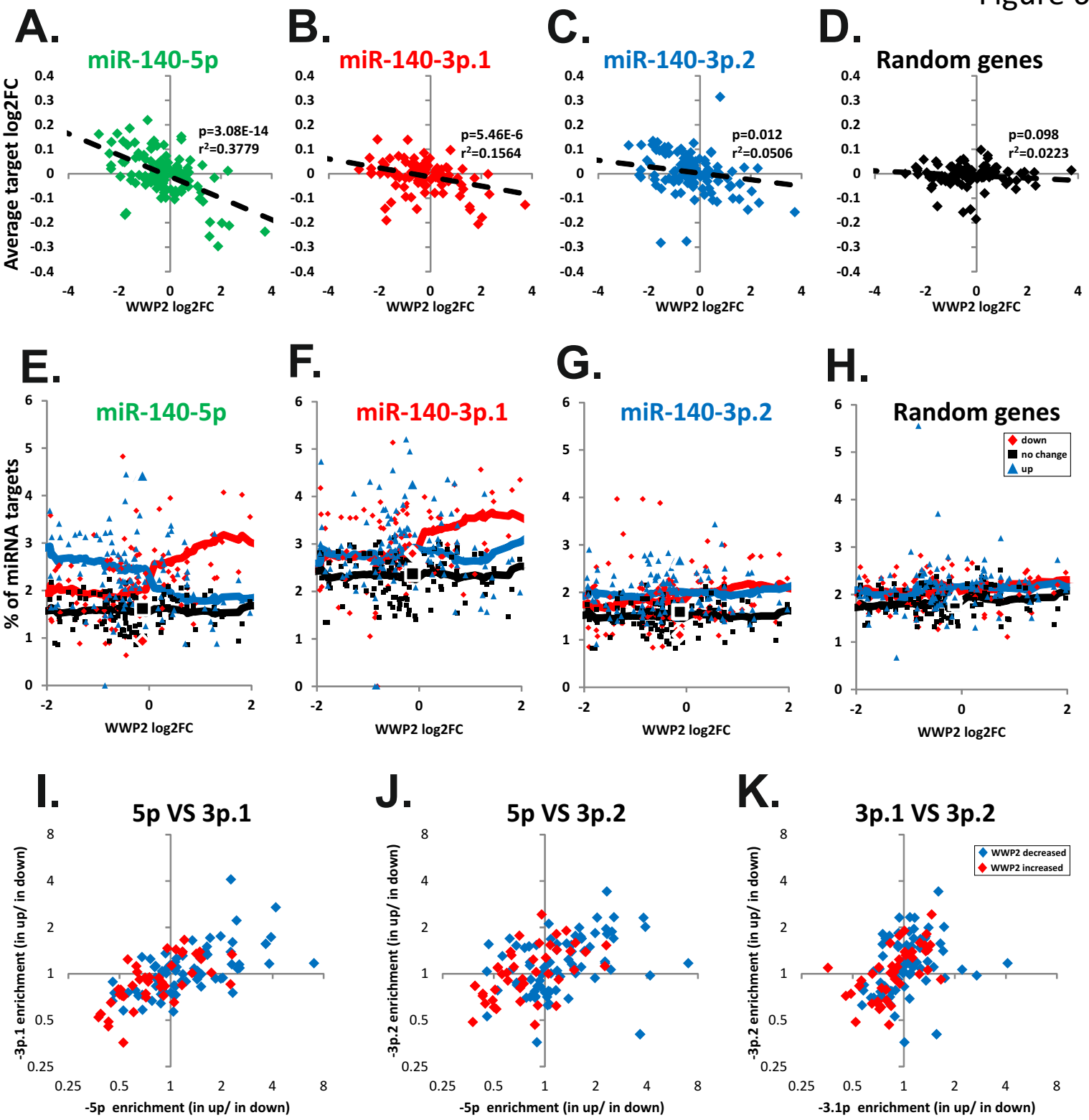

A.
Log2 FC
8
6
4
2
0
-2
-4
CD44
ENG (CD105)
THY1 (CD90)
NT5E (CD73)
SOX9
ACAN
COL2A1
MATN3
B.
Average Log2 FC
0
-0.05
-0.1
-0.15
-0.2
-0.25
-0.3
-0.35
miR-140-5p
miR-140-3p.1
miR-140-3p.2
only miR-140-3p.1
only miR-140-3p.2
all genes
C.
Enrichment (log10)
6
5
4
3
2
1
0
-1
-2
miR-140-5p
miR-140-3p.2
miR-140-3p.1
down
up

Figure 6


A.
miR-140-5p
p=3.08E-14
r²=0.3779
Average target log2FC
WWP2 log2FC
B.
miR-140-3p.1
p=5.46E-6
r²=0.1564
WWP2 log2FC
C.
miR-140-3p.2
p=0.012
r²=0.0506
WWP2 log2FC
D.
Random genes
p=0.098
r²=0.0223
WWP2 log2FC
E.
miR-140-5p
% of miRNA targets
WWP2 log2FC
F.
miR-140-3p.1
WWP2 log2FC
G.
miR-140-3p.2
WWP2 log2FC
H.
Random genes
down
no change
up
WWP2 log2FC
I.
5p VS 3p.1
-3p.1 enrichment (in up/ in down)
-5p enrichment (in up/ in down)
J.
5p VS 3p.2
-3p.2 enrichment (in up/ in down)
-5p enrichment (in up/ in down)
K.
3p.1 VS 3p.2
WWP2 decreased
WWP2 increased
-3p.2 enrichment (in up/ in down)
-3.1p enrichment (in up/ in down)