SCIENTIFIC REPORTS

OPEN

SUBJECT AREAS:
MIRNAS
RNAI

Received
13 May 2014

Accepted
9 July 2014

Published
8 August 2014

Correspondence and requests for materials should be addressed to I.R. (isidore.rigoutsos@jefferson.edu)

* These authors contributed equally to this work.

† Current address: Department of Pathology & Laboratory Medicine, The Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA.

Argonaute CLIP-Seq reveals miRNA targetome diversity across tissue types

Peter M. Clark[1]*†, Phillipe Loher[1]*, Kevin Quann[1], Jonathan Brody[2], Eric R. Londin[1] & Isidore Rigoutsos[1]

[1]Computational Medicine Center, Thomas Jefferson University, Philadelphia, PA, 19107, USA, [2]Department of Surgery, Thomas Jefferson University, Philadelphia, PA, 19107, USA.

To date, analyses of individual targets have provided evidence of a miRNA targetome that extends beyond the boundaries of messenger RNAs (mRNAs) and can involve non-Watson-Crick base pairing in the miRNA seed region. Here we report our findings from analyzing 34 Argonaute HITS-CLIP datasets from several human and mouse cell types. Investigation of the *architectural* (i.e. bulge vs. contiguous pairs) and *sequence* (Watson-Crick vs. G:U pairs) preferences for human and mouse miRNAs revealed that many heteroduplexes are "non-canonical" i.e. their seed region comprises G:U and bulge combinations. The genomic distribution of miRNA targets differed distinctly across cell types but remained congruent across biological replicates of the same cell type. For some cell types intergenic and intronic targets were more frequent whereas in other cell types mRNA targets prevailed. The findings suggest an *expanded* model of miRNA targeting that is more frequent than the standard model currently in use. Lastly, our analyses of data from different cell types and laboratories revealed consistent Ago-loaded miRNA profiles across replicates whereas, unexpectedly, the Ago-loaded targets exhibited a much more dynamic behavior across biological replicates.

MiRNAs are short non-coding RNAs (ncRNAs) that regulate their target mRNAs in a sequence-dependent manner thereby regulating the expression of the corresponding protein-coding gene[1]. MiRNAs are the best-studied group of ncRNAs and have been shown to be critical for many biological processes[2–5] and cancers[6,7], while exhibiting tissue and cell-state dependent expression profiles[8]. Ever since the first reported animal heteroduplex[2,3], lin-4:lin-14, it has been clear that a portion of the 5′ region of a miRNA plays a central role in the recognition of the miRNA's target. This portion typically spans positions 2–7 from the miRNA's 5′ end and is known as the 'seed.' The presence of the seed sequence's *reverse complement* (i.e. of contiguous Watson-Crick base pairing in the seed region), the localization in the 3′ untranslated region (3′UTR) of a messenger RNA (mRNA) and, occasionally, the conservation of a candidate sequence across genomes have been typical criteria for determining mRNA targets[1,2,4,9]. In addition to contiguous Watson-Crick base pairing in the seed region, non-standard interactions where the base pairing was interrupted by bulges have also been reported[2,4,5,10–19]. Analogously, other reports showed instances of "seed-less" interactions[5,15,16,20–29], targets located outside the 3′UTR[5,22,27,28,30–33], and targets that were not conserved amongst various species[5,15,33,34]. However, the prevalence of such non-standard interactions as compared to those that are anticipated by the standard model remains unclear.

The advent of CLIP-seq (cross-linked immunoprecipitation followed by next generation sequencing) techniques such as HITS-CLIP[35], PAR-CLIP[36], and iCLIP[37] has helped make great strides towards solving the problem of identifying miRNA targets with higher confidence. Rigorously speaking, CLIP-seq can identify miRNAs and targets that are part of the Ago silencing complex but does not directly establish *which miRNA* forms a heteroduplex with *which target*; the recently published CLASH[38] is a first attempt towards solving this problem biochemically. Nonetheless, determining the specifics of the heteroduplexes captured in AGO CLIP-seq experiments is possible through additional analysis. Indeed, several such methods have already been developed by others[36,39–48] as well as by us[14].

Continuing our earlier work with non-standard heteroduplexes[5,15–17,26,49] we expanded on our previously reported CLIP-seq analysis method[14] and used it to investigate the *sequence* (i.e. possible presence of one or more G:U pairs) and *architectural* (i.e. possible presence of a bulge on either the miRNA or the target side) preferences that are present in the seed region of miRNA:target heteroduplexes. The result is a very large collection of computationally predicted interactions across the genome that are derived from seven different cell sources and two organisms. Our analyses included public datasets and CLIP-seq datasets generated in our laboratory from the hTERT-HPNE and MIA PaCa-2 cell lines.

Results

We analyzed a total of 34 Ago CLIP-seq datasets (four human and 30 mouse – Supp. Table 1). As has been pointed out previously[50] the HITS-CLIP and PAR-CLIP methodologies generate essentially the same results, an observation we were also able to recapitulate using public samples for which both types of data were available (see Supp. Table 2). In light of this and to ensure uniformity across the processed samples, we limited our analysis to public Argonaute HITS-CLIP (CLIP-seq) datasets only. We follow the approach that we published previously[14] for analyzing CLIP-seq datasets (CLIPSim-MC) and which is summarized in Figure 1 (see also Materials and Methods).

The analyzed biological replicates show congruence in the Ago-loaded miRNAs but not in the Ago-loaded targets. It is important to stress that in this sub-section we aim to address two important questions. First: are the profiles of the top-expressed, Ago-loaded miRNAs concordant across biological replicates from the same cell type/tissue? Second: are the profiles of the statistically significant Ago-loaded targets concordant across biological replicates from the same cell type/tissue? In other words, we simply inspect the Ago-bound RNA across biological replicates from the same tissue/cell type to determine the extent to which the miRNA:target heteroduplexes remain unchanged. These two questions are of immediate relevance in light of recent data[38,51] that suggest the possibility of a dynamic miRNA targetome. In what follows, we use the term "MRE cluster" to refer to genomic segments that do not correspond to any annotated miRNA locus and are delineated by a collection of overlapping reads (see also Methods for a detailed definition of the terms MRE, MRE motif, and MRE cluster). Each MRE cluster comprises *at least one* 'miRNA response element' (MRE) and typically encompasses a multitude of distinct and potentially overlapping miRNA binding sites for different miRNAs.

With regard to the first of these two questions, we find that the most abundant endogenous miRNAs that are loaded on Ago show a high degree of overlap across the biological replicates of a given cell type. Figure 2a shows the Spearman correlations amongst top-expressed miRNA for the analyzed datasets for which biological replicates were available. For all represented cell types, there is a high degree of correlation among the replicates for the top-expressed Ago-loaded miRNAs, indicating a very high consistency in the profile of top-expressed, Ago-loaded miRNAs within these datasets. This concordance is particularly striking in the pairwise comparisons of the replicates from the mouse CD4+ T-cell samples and for all 12 wild type (WT) and 12 miR-155 knockout (KO) samples.

With regard to the second of these two questions, we find that the concordance of the Ago-loaded miRNAs among the replicates does not extend to the MRE clusters. Our results indicate that the Ago-loaded MREs with statistically significant coverage have little overlap across biological replicates (Figure 2b). Additionally, and for each available tissue type in turn, we calculated the positional overlap among *all expressed* MRE clusters across the biological replicates (Figure 2c). We also calculated this overlap by restricting ourselves to only the *significantly expressed* MRE clusters (Figure 2d). The point of this exercise was to evaluate the extent of overlap exhibited by the MRE clusters in the biological replicates. In the ideal scenario, the same exact MRE clusters should arise in each biological replicate; however, as Figures 2c and 2d show, this is not the case. We report our calculations in terms of "the number of unique genomic positions that are captured by those MRE clusters and are present in at least *n* of the biological replicates available for the tissue or cell type at

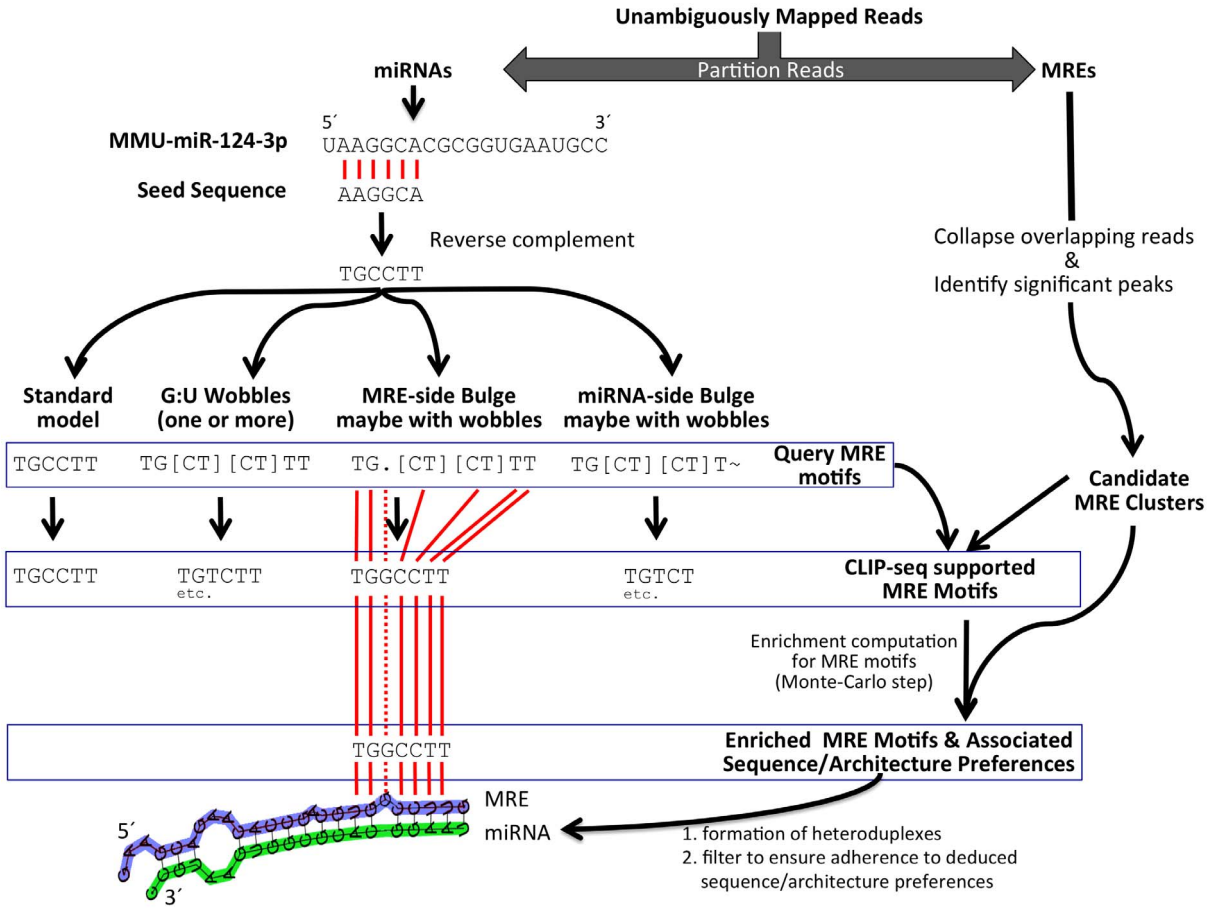

Figure 1 | Conceptual workflow of CLIPSim-MC. This schematic depicts the generation of possible expanded-model seed region formations and a CLIP-supported MRE-motif representing a single bulge on the MRE side of the resulting heteroduplex for mouse miR-124-3p.

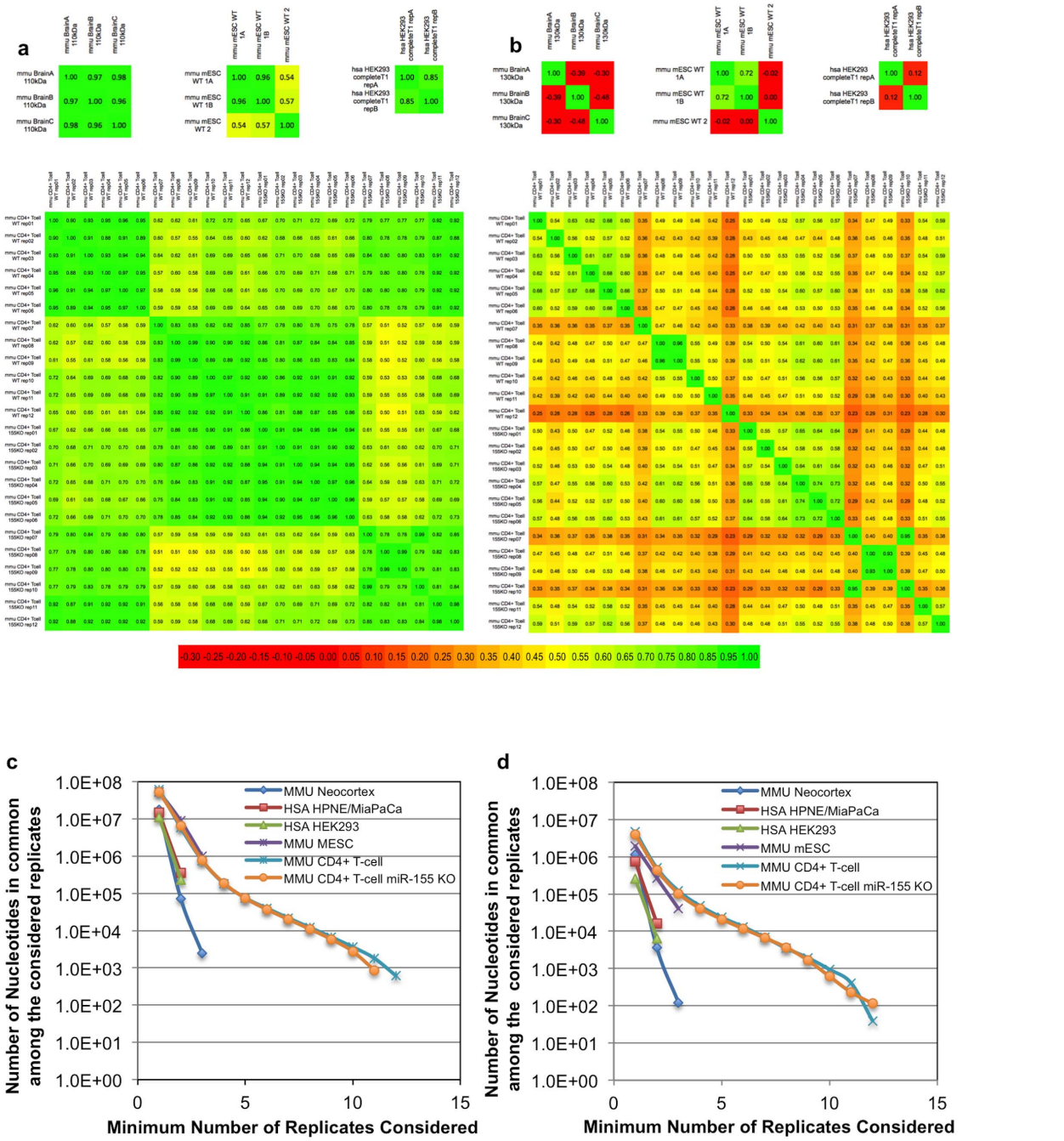

Figure 2 | Examining whether the top expressed miRNAs and MRE clusters recur in the biological replicates of each study. (a) Spearman correlation across biological replicates for the top-abundant miRNAs. Top row: mouse brain, mouse ESC, and human HEK293. Bottom row: 12 wild type (WT) and 12 miR-155 knockout (155 KO) CD4+ T-cells. (b) Spearman correlation across biological replicates for the statistically significant MRE clusters. Top and bottom rows are as in panel (a) above. (c) Number of unique genomic positions that are captured by all MRE clusters and are common to at least *n* of the biological replicates available for the tissue or cell type at hand (the value of *n* is shown on the X-axis). This effectively gauges the degree of recurrence of a specific miRNA target at a specific genomic position (represented by the MRE cluster) across the available biological replicates. (d) In this panel we repeat the calculations of panel (c) considering only the statistically significant MRE clusters in each biological replicate. Note: even though the hTERT-HPNE/MIA PaCa-2 curve does not correspond to biological replicates but to two distinct cell types from the same tissue (pancreas) we include it in panels (c) and (d) for comparison purposes.

hand." In Figure 2d we restrict the calculation to using statistically significant MRE clusters only. Clearly, the value of 'n' ranges from 1 to the total number of available replicates. Our results show a ten-fold decrease in the number of bases covered by at least two replicates compared to the number of bases covered by at least one replicate. Moreover, we note that the imposition of the statistical significance constraint alone reduces the breadth of genomic coverage by ~10 fold. As we increase the minimum number of replicates in which an MRE cluster is required to occur, the number of bases spanned by the surviving MRE clusters decreases exponentially underlining a dynamic nature in the targeted MREs.

The high correlation that we observed with the miRNA component of the Ago-loaded miRNA:MRE heteroduplexes indicates that the lack of correlation among Ago-loaded MREs does not reflect a

Figure 3 | Distribution of MRE clusters. Distribution of statistically significant MRE clusters (p-value ≤ 0.05) across intergenic, intronic and exonic space is shown separately for each sample.

technical issue but rather suggests the existence of a miRNA target repertoire that is highly dynamic and transient in nature, an observation recently reported by others as well[38,51]. Our finding is further supported by the fact that the replicates show limited Ago-target footprint overlap even when no statistical filtering is applied. An alternative explanation could be a possible dependence of the targeted transcript populations on cell cycle. In light of this observation and given the diversity in the breadth and depth of coverage across replicates, we chose to analyze and apply statistical significance filtering separately to each replicate: had we required that an MRE be present in two or more of the replicates we would have restricted our focus to an artificially small number of bases (evidenced by Figures 2c and 2d) neglecting the information that results from the apparently dynamic nature of the miRNA targetome.

The MRE clusters are spread across all genomic regions. Our analyses reveal that, for most of the analyzed samples, a considerable portion of the statistically significant (p-value ≤ 0.05) MRE clusters are located beyond the exonic space. Indeed, the *intergenic* portion of the statistically significant MRE clusters ranges between 10 and 25%. The HEK293 samples are an exception with ~45% of the MREs being intergenic (Figure 3). Looking at the data across samples, we find several of the intergenic MRE clusters in lncRNAs (human: 611 mouse: 4,107) and pseudogenes (human: 199 mouse: 2,284) – see Supp. Figures 3 and 4.

The analyzed datasets exhibit a wider variation in their portions of *intronic* and *exonic* MREs. In the mouse embryonic stem cell (mESC) samples, a mere 5% of the statistically significant MRE clusters are found in exonic space whereas the majority (~75%) arise from intronic loci. The mouse CD4$^+$ T-cell samples exhibit the opposite behavior: here, the majority (~80%) of the statistically significant MRE clusters derive from exonic space; an additional ~18.5% derive from intergenic space and the remaining ~1.5% from intronic loci. Lastly, the mouse brain replicates, similarly to the mESC ones, exhibit a notable abundance (70%) of intronic MRE clusters: the remaining MREs are evenly divided among exonic and intergenic space. Figure 3 also makes evident that across samples, the majority of exonic MREs arise from 3′UTRs, with coding sequences (CDS) contributing the second highest number of MREs. Lastly, it is important to note that despite the diversity of the MRE clusters among biological replicates (Figure 2b, 2c, and 2d), the replicates exhibit far greater similarity with regard to the subset of the genomic space (i.e. intergenic, intronic, exonic-5′UTR, exonic-CDS, exonic-3′UTR) where the MREs are found (Figure 3).

Most MRE loci can be unambiguously associated with a single miRNA. For each dataset, we considered further only enriched (FDR ≤ 0.05) MRE-motifs (and associated miRNA informed heteroduplex architectures). As described in Methods, we only kept those of the enriched miRNA:MRE-motif pairs, derived from CLIPSim-MC, for which the associated heteroduplex exhibits bonded base pairs beyond the seed region and the RNA folding matches the prescribed architecture from which the MRE-motif was originally derived (Figure 1). As shown in Figure 4, and across all studied datasets, we can unambiguously identify the miRNA participating in a miRNA:MRE heteroduplex for ~70% of all heteroduplexes: i.e. in these cases, the MRE locus is paired with a single targeting miRNA. For an additional ~20% of the formed miRNA:target heteroduplexes the MRE locus is paired with exactly 2 targeting miRNAs. We *manually* examined the instances where an MRE is paired-up with two or more miRNAs and invariably found that in such cases the targeting miRNAs are paralogous members of

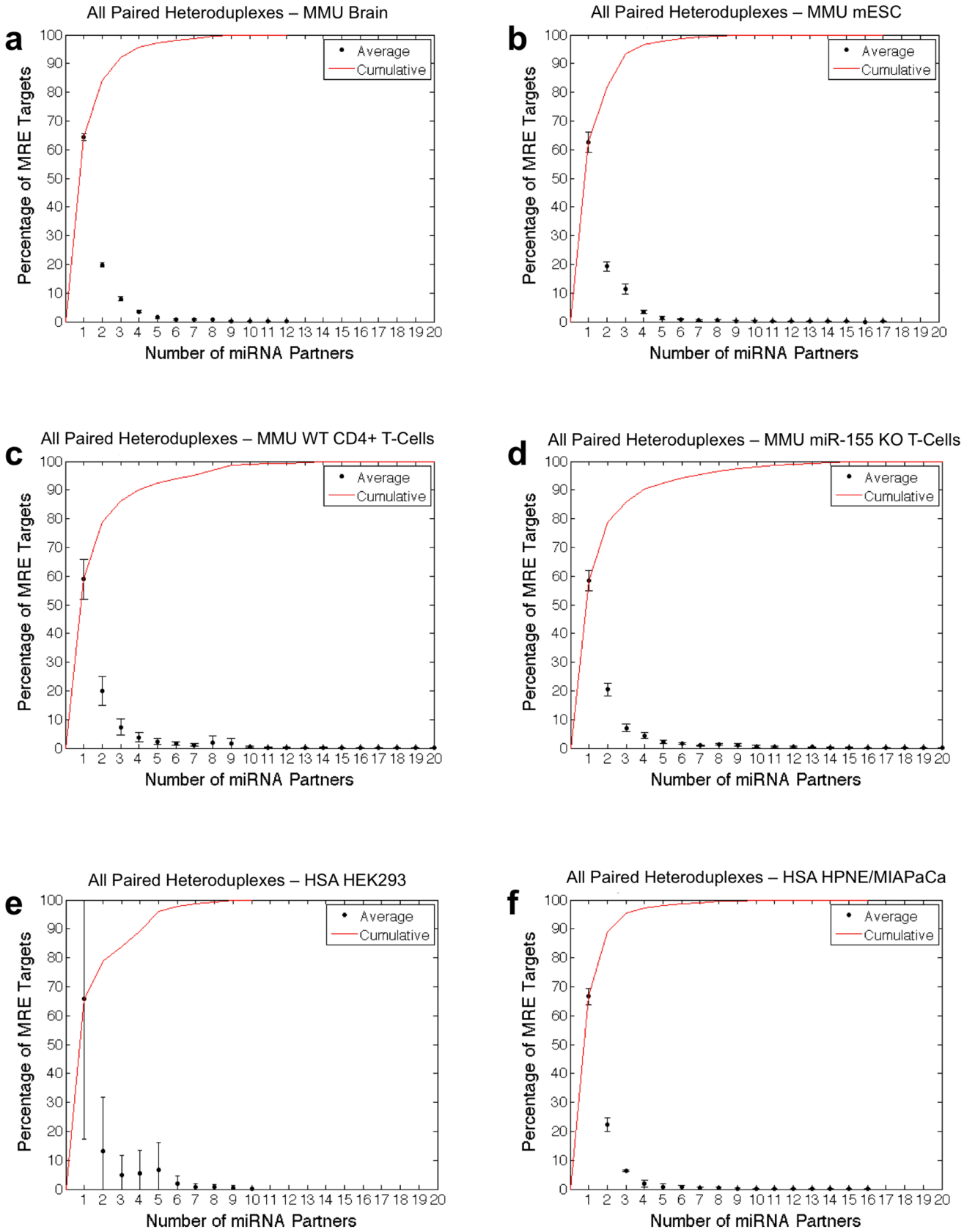

Figure 4 | Distribution of paired heteroduplexes. Panels (a) through (f) show the number of distinct miRNAs associated with a given MRE locus. Data points represent the average over all the replicates of the corresponding sample. Error bars represent the standard deviation across the replicates. The MRE motifs of all considered heteroduplexes have an FDR ≤ 0.5. For ~70% of all miRNA:MRE heteroduplexes we can unambiguously identify a single miRNA for a given MRE.

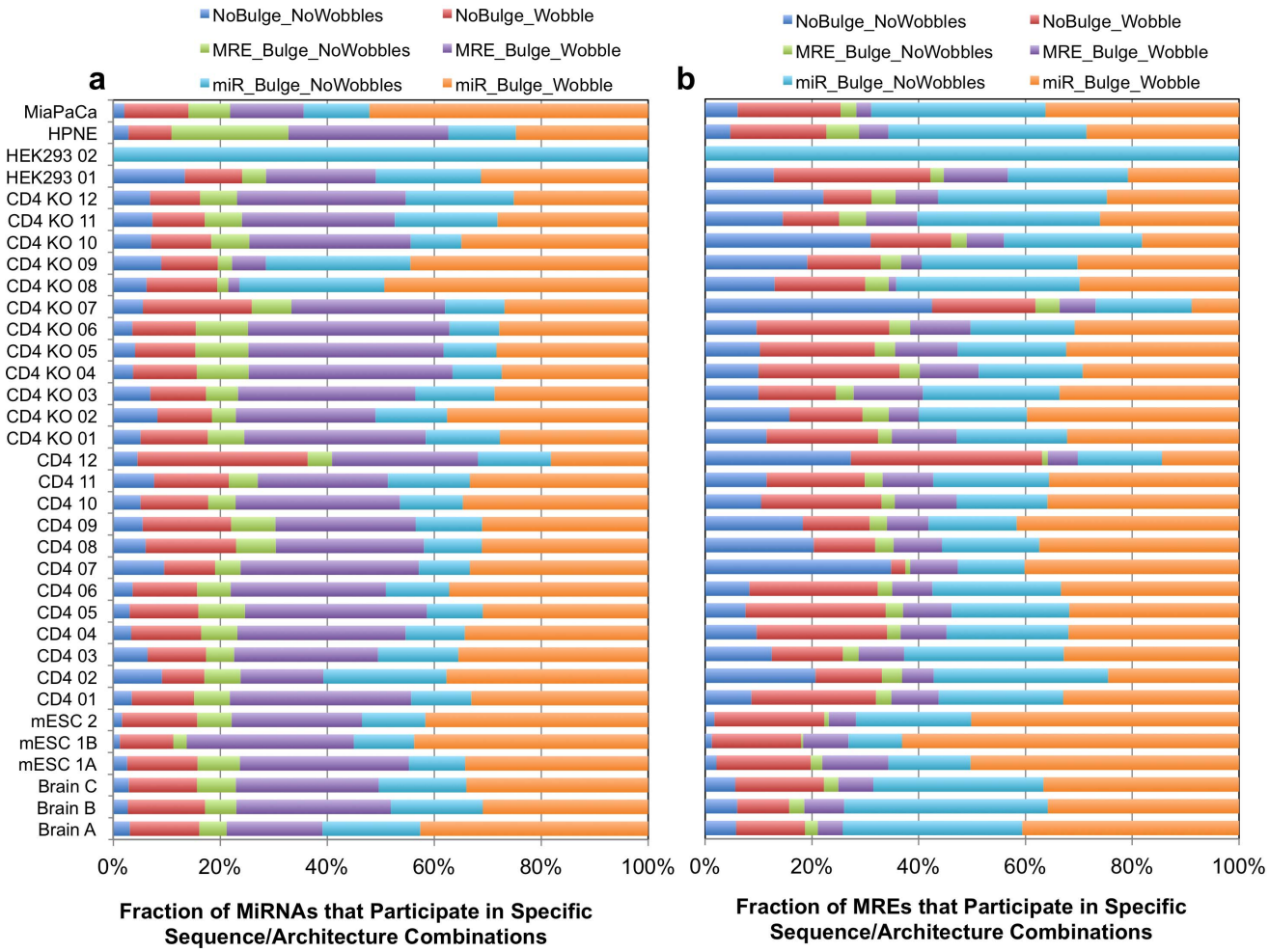

Figure 5 | Distribution of the specific sequence/architecture choices for the seed region among the derived heteroduplexes. (a) The distribution as seen from the standpoint of miRNAs and separately for each analyzed sample. E.g, according to the shown results, more than 50% of the endogenous miRNAs in MIA PaCa-2 form heteroduplexes with a miRNA-side bulge and at least one wobble in the seed region. (b) The distribution as seen from the standpoint of MREs and separately for each analyzed sample. E.g., according to the shown results, more than 35% of the MIA PaCa-2 MRE's participating in heteroduplexes have a miRNA-side bulge and at least one wobble in the seed region.

miRNA families with known and extensive sequence similarity (e.g. let-7a/b/c/…, miR-29a/b/c, miR-103/107, etc.). This ambiguity is inherent and anticipated given the high sequence similarity among the paralogous members of these miRNA families. In order to generate conservative estimates, for the remainder of our analysis, we will work with only the miRNA:MRE pairs for which the MRE locus is paired with a *single* endogenous miRNA.

Both standard- and expanded-model miRNA:target heteroduplexes are frequent. Figure 5 shows, for each analyzed dataset, the breakdown of the enriched miRNA specific sequence/architecture arrangements (FDR $\leq$ 0.05) for the heteroduplexes in which an MRE is targeted by a single miRNA. In particular, Figure 5a shows for each sample what fraction of the endogenous miRNAs form heteroduplexes with the shown sequence/architecture arrangement in the seed region. Analogously, Figure 5b shows for each sample what fraction of the MRE loci form heteroduplexes with the shown seed/architecture arrangement in the seed region. Taken together, these two plots highlight the following observation for the final set of heteroduplexes for a given sample: although a specific sequence/architecture choice for the seed region for a miRNA may be proportionally small, this particular choice may be used in forming heteroduplexes with a proportionally larger pool of MREs within the sample. Proportionally, and across all analyzed datasets, the standard model (non-bulged contiguous Watson-Crick base-pairing in the seed region) represents the least abundant category (~3–12%). The most abundant category, in all datasets, corresponds to G:U wobbles in the seed region together with a single miRNA bulge within the seed (30–50%). The second most abundant category comprises G:U wobbles in the seed region and a single bulge on the MRE side of the seed region (20–40%). Non-bulged heteroduplexes with at least one G:U wobble in the seed region account for an additional 15–25% of the cases. Examination of admissible formations that contained a bulge on either the miRNA or the MRE side of the seed region revealed that individual miRNAs have distinct bulge-positioning preferences. However, when we considered all of the heteroduplexes containing a bulge within the seed region of the heteroduplex that survived our analyses we found that bulges were equally likely at all seed region positions of either the miRNA or the MRE. It is evident that despite the abundance of heteroduplexes containing a bulge within the MRE and at least one G:U wobble, these heteroduplexes correspond to proportionally fewer MREs when compared to the other heteroduplex architectures that we considered.

In concordance with previous reports the 3′UTRs harbor a large portion of all the exonic loci in almost all datasets. In analogy with Figure 5, we examined the distribution of elucidated architectures for 3′UTRs (as well as 5′UTRs and CDSs) and found

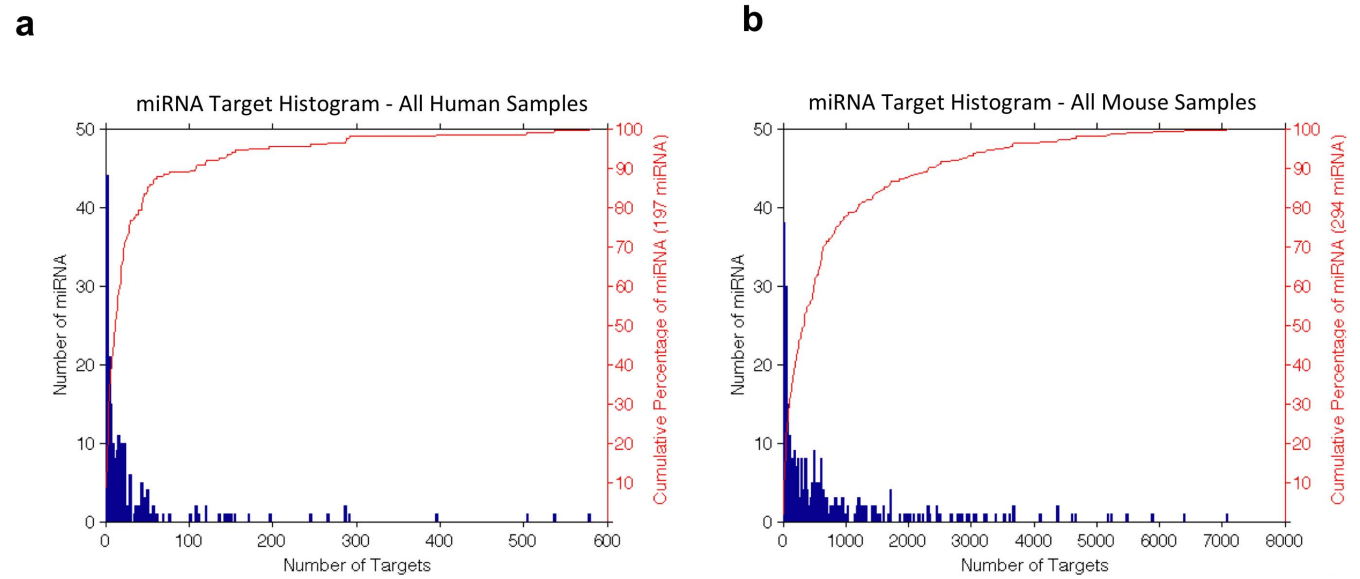

Figure 6 | Distribution of the number of endogenous miRNAs that are associated with a given number of distinct targets. (a) Human data. (b) Mouse data. The secondary Y-axis shows the cumulative distribution of the expressed miRNAs that are associated with a given number of targets. Only targets i.e. MREs that are associated with a single miRNA were considered in this calculation.

a prevalence of expanded-model heteroduplexes – see Supp. Figures 5 and 6.

Overlap with other available predictions. Two recent publications studied in detail two mouse miRNAs, miR-124[52] and miR-155[53], and reported findings on the miRNAs' targeting preferences. These two miRNAs are good test cases as their preferences represent instances of non-standard interactions. In the case of miR-124, only *exonic* MRE clusters were considered in the original report[52], thus we repeated our CLIPSim-MC simulations for the mouse brain samples by considering only the *exonic* subset of all MRE clusters instead of genome-wide MRE clusters. As can be seen from Table 1A, the guanine bulge site at seed position 6 on the MRE side that was reported[52] is correctly captured through the enriched TGGCCTT MRE-motif. The MRE-motif corresponding to the standard model of targeting, i.e. the one whose seed region sequence is the reverse complement of miR-124-3p's seed is also among the statistically significant ones as are several additional MRE-motifs capturing expanded-model interactions (Table 1A). The complete list of *exonic* preferences (sequence and architecture) for miR-124-3p and the corresponding MREs are available on line at: https://cm.jefferson.edu/tools_and_downloads/clip_2014/output_exonic/mmu_miR_124_3p.output_exonic_bymiR.txt.

For the mouse miRNA miR-155-5p, all of the enriched MRE-motifs and corresponding formations that result from our analysis are presented in Table 1B. The entries of Table 1B show that in addition to the MRE-motif (GCATTA) that corresponds to the standard model and is enriched across many replicates, we find additional enriched expanded-model formations including several of those previously reported[53].

We also performed functional enrichment of miR-155 targets identified by our analysis in order to determine which gene ontology (GO) biological process terms are enriched among the identified mRNA targets of miR-155. The analyses were carried out using DAVID[54,55] and only those biological processes with an FDR $\leq$ 0.05 were considered further (Supp. Table 3). Our results indicate that miR-155 targets mRNAs significantly involved in transcriptional regulation, cell fate and differentiation as well as several other immune related processes. Our results are consistent with previous CLIP-seq based findings[53], and with the relevant T-cell literature[56,57].

Of the available public repositories of CLIP-seq analyzed data[47,58], Starbase[47] makes their predictions available in a manner that permits direct comparisons. Starbase contains 601,189 human and 111,809 mouse target predictions with ~93% of these predictions being located in 3′UTR space. We note here that several of the target-site prediction algorithms that Starbase uses report only standard-model targets, which leads to an over-representation of such formations in the Starbase pool of data. Consequently, for this comparison, we focused on the 3′UTR and canonical subset of our predictions. We find that 76.5% of our human standard-model predictions (2,355) and 42.3% of our mouse standard-model predictions (12,047) are identical to those reported in Starbase (Supp. Table 4). The difference is due to the fact that Starbase reports many more human targets, and we report many more mouse targets due to the specifics of the samples analyzed: note that for the mouse genome, we report nearly 5 times as many statistically significant heteroduplexes as we do for our human predictions.

MiRNAs can have many distinct targets in a given cell type. The analyzed datasets were obtained from diverse sources and allow us to shed some light on the question of how many mRNAs are targeted by a miRNA. We re-emphasize that in what follows, we focus only on MREs for which we can identify a single targeting miRNA within the corresponding dataset and as such our estimates are conservative and represent *lower bounds* of the true number of targets that a miRNA can have.

After processing each of the datasets separately, we formed the union of these miRNA:MRE interactions across the replicates of each cell type. From the pooled set of data, we find that a notable portion of the top-expressing miRNAs have hundreds of distinct targets each. As shown in Supp. Fig. 1 the magnitude of the miRNA targetome differs across the five tissues that we analyzed. In mouse brain, ~40% of the 106 top-expressing miRNAs have at least 55 distinct targets each; in mESC, ~40% of the 165 top-expressing miRNAs have at least 140 distinct targets each; in wild-type mouse $CD4^+$ T-cells, ~40% of the 177 top-expressing miRNAs have at least 415 distinct targets each; in mmu-miR-155 KO Cd4+ T-cell samples, ~40% of the 164 top-expressing miRNAs have at least 200 distinct targets

Table 1 | Enriched seed-region formations. a) Enriched seed-region formations involving MMU-miR-124-3p and their associated FDR values. The results correspond to using only exonic MRE clusters from the three mouse brain datasets. b) Enriched seed-region formations involving MMU-miR-155-5p and their associated FDR values. The results were obtained by considering MRE clusters across the genome from 12 CD4$^+$ T-cell mouse datasets. For each seed-region formation variant P-values were calculated by fitting the expected count distribution of the variant with a negative binomial distribution followed by multiple test correction using the Benjamini-Hochberg procedure

a Enriched Exonic MRE Motifs in Mouse Brain Replicates (MMU miR-124-3p) and FDR values

Architecture	Architecture Description	MRE Motif	Median FDR	Enriched Within *n* Replicates
~G[CT][CT]TT	miR_Bulge_NoWobbles	GCCTT	2.80E-03	3
TG[CT][CT]T~	miR_Bulge_NoWobbles	TGCCT	1.40E-04	3
TG[CT][CT]T~	miR_Bulge_Wobble	TGTCT	6.85E-03	2
T.G[CT][CT]TT	MRE_Bulge_NoWobbles	TTGCCTT	3.33E-03	2
TG.[CT][CT]TT	MRE_Bulge_NoWobbles	TGGCCTT	2.10E-02	1
TG[CT][CT]TT	NoBulge_NoWobbles	TGCCTT	1.00E-06	3

b Enriched MRE Motifs in CD4$^+$ T-Cell Replicates (MMU miR-155-5p) and FDR values

Architecture	Architecture Description	MRE Motif	Median FDR	Enriched Within *n* Replicates
G[CT].[GA]TT[GA]	MRE_Bulge_Wobble	GTTGTTG	7.30E-05	4
G[CT].[GA]TT[GA]	MRE_Bulge_Wobble	GCTGTTG	1.39E-02	5
G[CT][GA]TT[GA]	NoBulge_NoWobbles	GCATTA	1.09E-02	4
~[CT][GA]TT[GA]	miR_Bulge_Wobble	TGTTG	4.69E-03	4
G[CT][GA]TT[GA]	NoBulge_Wobble	GTGTTG	3.77E-02	3
G[CT].[GA]TT[GA]	MRE_Bulge_Wobble	GCTATTG	3.77E-02	1
G[CT][GA].TT[GA]	MRE_Bulge_Wobble	GTGCTTG	4.19E-02	1
~[CT][GA]TT[GA]	miR_Bulge_Wobble	CATTG	4.92E-02	1

each; in the hTERT-HPNE and MIA PaCa-2 cell lines, ~40% of the 160 top-expressing miRNAs have at least 15 distinct targets each; and, in HEK293 cells, ~40% of the 67 top-expressing miRNAs have more than 5 distinct targets each. These data, taken together with the results of Figures 3, suggest that each miRNA has a large repertoire of cell-type specific targets. The findings also indicate that a given endogenous miRNA can have a rather distinct targetome within a given tissue, with a given endogenous miRNA targeting many MREs in one tissue type and fewer in another. As an example let us consider miR-18a-3p, a member of the miR-17/92 oncogenic cluster that is conserved across vertebrates[59–61]. MiR-18a-3p is associated with 917 unique MREs across all mESC samples, 253 unique MREs across all mouse CD4$^+$ T-cells, 33 unique MREs across all miR-155 KO mouse CD4+ T-cells, and 25 MREs in the hTERT-HPNE/MIA PaCa-2 samples. On the other hand it is not associated with any targets within the mouse brain samples or in the HEK293 cell line samples.

To appreciate how many distinct MREs may be targeted by a single miRNA across *different tissues* of the *same organism* we formed the union of miRNA:MRE interactions we obtained from the three analyzed mouse cell types (30 datasets) and the two human cell types (4 samples corresponding to three cell lines) respectively. We only considered MREs with an unambiguously determined targeting miRNA across all mouse samples and find 228,688 unique MREs that are targeted by 294 unique miRNAs through 233,364 unique interactions. For the human samples, we find 7,851 unique MREs targeted by 197 unique miRNAs through 7,866 unique interactions. In Figure 6a, we consider only MREs that have been associated with a single miRNA in our analysis of human samples to derive the count of miRNAs (primary Y-axis) that are associated with a given number of predicted targets (X-axis). All analyzed human samples are considered for this purpose. The secondary Y-axis shows the cumulative distribution of the expressed human miRNAs that are associated with a given number of predicted targets. This histogram is meant to provide estimates for the number of distinct targets that an endogenous miRNA can have across tissues/cell types. Figure 6b shows the same histogram for the analyzed mouse tissues/cells. For the mouse datasets, more than 20% of the 294 analyzed miRNAs have more than 1,000 distinct targets each. These findings demonstrate that numerous discrete MRE loci are unambiguously associated with a putative targeting miRNA. In the Supplement, we also address and present results for a related question namely how many distinct MREs can a given miRNA target in an mRNA.

On-line exploration of the data. The complete data (in both miRNA-centric and genome-centric views) for each analyzed human and mouse miRNA for all 34 datasets are available for interactive exploration on-line at https://cm.jefferson.edu/clip_2014/. The data has been compiled in two different ways: First, we provide a *miRNA-centric* view: for each miRNA, we present the sequence of the targeted MRE-motif, and the corresponding p-value and FDR for each analyzed sample (one per sample). Also stated is the resulting MRE-formation, e.g. G:U wobbles, MRE-side bulge, miRNA-side bulge, etc. The second view is *genome-centric* and is meant to acknowledge the increasing realization that miRNAs target numerous transcripts, protein-coding as well as non-coding RNA. In this case, we list the genome identifier, chromosomal location of the strand where the MRE is found, cell type in which the interaction in encountered, identity of the replicate supporting the target, p-value and FDR for the MRE motif, identity of the targeting miRNA, and the Gibbs free energy of the associated heteroduplex. For those miRNAs that are not among the 34 analyzed CLIP-seq datasets, we make available version 2.0 of the rna22 method[17] at https://cm.jefferson.edu/rna22v2/.

Discussion

Through our analysis of 34 independent CLIP-seq samples, we identified computationally predicted, high confidence, statistically enriched seed-region formations and full-length heteroduplexes. With regard to the location of the miRNA targets our analysis shows that many statistically significant MREs are present in exonic space, which is expected, with the rest of them located in intergenic and intronic regions. The portion of exonic MREs was consistent across biological replicates while it ranged from sample to sample: from 20–40% in HEK293 cells and mouse brain to ~75% in mouse CD4+ T-cells. The three mESC datasets represented an exception to these findings in that nearly two thirds of the statistically significant

MREs were located in introns. Among the exonic MREs, approximately half were located in the 3′UTRs.

Additionally, we examined the specifics of the architecture (presence or absence of bulges) and sequence (presence or absence of G:U wobbles) preferences for the statistically significant heteroduplexes. In concordance with earlier findings, our analysis of these heteroduplexes revealed a biologically diverse miRNA targetome comprising MREs that participate in both standard and expanded seed-region formations with the targeting miRNA. The expanded formations include various combinations of G:U wobbles and single nucleotide bulges within the seed-region of the heteroduplex and outnumber the standard formations. Moreover, we found that many of the endogenous top-expressed miRNAs of a given sample exhibited concrete non-standard targeting preferences that were cell-type specific. Looking across all samples, approximately one third of the statistically significant MREs participated in standard seed-region interactions (contiguous Watson-Crick base pairing). Formations that involved a single bulge on the miRNA side of the heteroduplex as well as the presence of at least one G:U wobble represented another abundant category.

Another thing we considered was the profiles of Ago-loaded miRNA and targets. With regard to the top-abundant endogenous miRNAs, we found them to be consistently present across the replicates of a given sample. Somewhat surprisingly, the profiles of the MREs exhibited a more dynamic behavior across the replicates of the same cell type. Interestingly, the breakdown of MRE locations across intergenic-intronic-5′UTR-CDS-3′UTR space was preserved across the replicates even though the exact target locations were not. These observations held true for all considered cell/tissue types and for both human and mouse suggesting that a complex and dynamic process is at play. These results add to the growing evidence that a set of highly expressed miRNAs regulate a dynamic pool of MREs transcribed from across the genome[38,42,47,62–64].

Our findings also shed some light on the number of distinct transcripts that can be targeted by a miRNA. Indeed, we found evidence for a rich target repertoire for many miRNAs, a repertoire that can comprise hundreds of distinct targets for a given miRNA in the same cellular context. The number of distinct targets for a given miRNA increases further when one considers the miRNA's targetome across cell types.

We conclude by commenting on one more ramification of our results from the standpoint of miRNA-effected regulation. The apparent abundance of non-protein-coding miRNA targets in conjunction with the finding that several miRNAs can have many targets and that an mRNA can be targeted by many miRNAs simultaneously provides additional support to the concept of miRNA sequestration[65,66] and competing endogenous RNAs (ceRNAs)[67]. The diversity of involved genomic transcripts and the large number of promiscuous miRNAs encountered in each of the five cell types indicate that a large number of ways exist in which sequestering of miRNAs by sponges and ceRNAs through target decoying can regulate protein-coding transcripts.

Methods

Cell culture, Ago HITS-CLIP and RNA-sequencing. The hTERT-HPNE and MIA PaCa-2 cell lines were obtained from American Type Culture Collection (Manassas, VA) and from Dr. Jonathan Brody, and propagated in Dulbecco's Modified Eagle Medium supplemented with 10% fetal bovine serum and 1% penicillin/streptomycin (Cellgro, Manassas, VA). Ago HITS-CLIP was performed as described previously[35] with modifications to increase stringency[68]. Briefly, cells were grown to 70% confluency, washed once with PBS and UV irradiated at 254 nm for a total energy dispersion of 600 mJ/cm^2 (Spectroline, Westbury, NY). RNA digestion was carried out as per Hafner et al[36]. Cell lysates were treated initially with RNAse T1 at a concentration of 1 U/μl for 15 minutes at room temperature in PXL buffer prior to co-immunoprecipitation of RNA-protein complexes on protein A Dynabeads (Life Technologies) using the pan-Ago antibody 2A8 for 4 hours at 4°C (Millipore, Billerica, MA). Beads were then washed twice with PXL buffer and subjected to a secondary, complete RNA digestion with 100 U/μl of RNAse T1 for 15 minutes at room temperature. Following complete digestion, CLIP-RNAs were liberated from their on-bead protein complexes by treatment with 4 mg/ml proteinase K and subsequent phenol/chloroform extraction as described earlier[35]. CLIP-RNA libraries were constructed using the small RNA library preparation protocol as described above. All libraries were sequenced on Applied Biosystems 5500XL sequencers (Life Technologies).

Definitions of "MRE" and "MRE motif". The term *miRNA response element* or MRE was originally coined to capture the full span of a miRNA target[69] and not just the target's six-nucleotide-long seed region. Since then, the term been overloaded and is also used to refer to the target's nucleotide stretch that interacts with the seed region of the targeting miRNA. In what follows, we use the more specific term "MRE-motif" to refer to the portion of the target opposite the miRNA's seed region (positions 2–7 inclusive) and use "MRE" to refer to the full-length miRNA target. Also, we will use the term "formation" to refer to an arrangement of the base pairs in the seed region that comprises any combination of sequence (Watson-Crick pairs or G:U wobbles) and architecture (bulge or no bulge). Finally, we use the term "heteroduplex" to refer to miRNA:target interactions that span the *full length* of the targeting miRNA (as opposed to only the seed region). In all of our analyses, we use the string of the MRE as a reference string; as such we need to introduce notation that will allow us to indicate the presence and location of bulges on either the MRE or the miRNA side, and of G:U wobbles. To this end, we use a '.' to denote a seed-region bulge on the side of the MRE-motif (target). For example, TG.CCTT indicates that the nucleotide of the MRE-motif that occupies the '.' position, e.g. $\underline{G}$ in 5p → TG$\underline{G}$CCTT → 3p, will be unpaired. Analogously, we use a '∼' to denote a seed-region bulge on the side of the miRNA. For example, TG∼CTT indicates that the nucleotide of the miRNA that occupies the position across the '∼' symbol, e.g. $\underline{G}$ in 3p ← AC$\underline{G}$GAA ← 5p, will be unpaired. To denote the potential of G:U wobbles forming we use bracketed expression: e.g. the last four positions of 5p → TG$\underline{G}$[CT][CT][CT][CT] → 3p.

Preprocessing of raw reads and sequence mapping. In addition to our in house samples, we also analyzed 32 publicly available CLIP-seq samples that were precipitated using monoclonal antibodies against Argonaute 2 from four distinct studies that represent four cellular phenotypes[35,50,52,70]. Following adapter sequence removal and quality trimming with the help of cutadapt[71], reads were mapped to their respective reference genome (human-hg19, mouse-NCBIM37) using SHRIMP2[72]. Only reads that could be placed unambiguously on the genome by allowing up to 4% mismatches (replacements only – no insertions or deletions were permitted) were considered in the subsequent analyses (Supp. Tab. 1).

Selecting miRNAs and MREs. We used the reads that mapped to the mature miRNA sequences (human and mouse) listed in Rel. 20 of miRBase[73] to generate endogenous miRNA profiles for each analyzed CLIP-seq sample. We identified the top-expressed miRNAs on a per sample basis by keeping only those miRNAs with abundance that was within 10 PCR cycles (a ratio of 1 : 1024) relatively to the sample's most abundant miRNA. Unambiguously-mapped reads that did not map to miRNA loci were taken to pinpoint MREs and were merged into "MRE clusters." MRE clusters are thus defined by overlapping reads that do not map to any annotated miRNA locus and may contain multiple target sites for a variety of miRNAs. We required that each MRE cluster comprise a minimum number of overlapping reads before it could be selected for subsequent analysis: this minimum required number of reads is determined by adapting a previously reported method[74] and carried out in a sample-specific manner that takes into account the depth of sequencing. Only statistically significant MREs (p-value $\leq$ 0.05) were kept for further processing. Considering the reported time-dependence of miRNA-targeting among biological replicates[38,42,62] and in order to be comprehensive in our characterization of the analyzed samples we identified and analyzed MRE clusters *separately* for each sample (see also Results, Figure 2, and Supp. Table 1). For the three mouse brain samples[35], MRE clusters were formed from the 130 kDa sample set only. The remaining CLIP-seq datasets included three biological replicates from mouse embryonic stem cells (mESCs)[70], 12 wild-type replicates 12 miR-155 knockout (KO) from mouse CD4$^+$ T-cell samples[53], two biological replicates from human embryonic kidney (HEK293) cells[50], and two CLIP-seq datasets that we generated from the hTERT-HPNE and MIA PaCa-2 cell lines (SRP034075).

Enumerating standard- and expanded-model seed-region formations. For each endogenous miRNA expressed in a given sample, we enumerated the following putative MRE-motif variants: a) the exact reverse complement of the miRNA's 6-nt seed region (this is the standard-model MRE-motif); b) all possible variants of the reverse complement that would necessitate that one or more G:U wobble base pairings, but no bulge, be formed if the corresponding heteroduplex were realized; c) all possible variants of the reverse complement that would require *a single bulge* on the miRNA side, but no G:U wobbles, if the corresponding heteroduplex were realized; d) all possible variants of the reverse complement that would require *a single bulge* on the MRE side, but no G:U wobbles, if the corresponding heteroduplex were realized; e) all possible variants of the reverse complement that would facilitate a single bulge on the MRE side of the potential heteroduplex with at least one G:U wobble within the seed; and f) all possible variants of the reverse complement that would facilitate a single bulge on the miRNA side of the putative heteroduplex in combination with at least one G:U wobble base pair within the seed region (Figure 1). In the presence of a single-nucleotide bulge, the MRE-motif will span five nucleotides (if the bulge is on the miRNA side) or seven nucleotides (if the bulge is on the target side). Because of this enumeration, these candidate formations include both standard-model and

expanded-model arrangements; also, because of the way our method arrives at these candidates we obviate any biases that could have been introduced by the use of target prediction tools to generate miRNA:target candidates from CLIP-seq data[35,42–44,46–48,50,75,76]. We refer to heteroduplexes that fall in cases b) through f) inclusive as instances of an "expanded model" of miRNA targeting.

Statistical enrichment of seed-region formations (CLIPSim-MC). The observed counts for each observed seed-region formation were calculated by finding the number of instances of the variant within the pool of MRE clusters. The expected count distribution for each observed seed-region formation was determined by carrying out a Monte-Carlo simulation in which each observed MRE-motif is queried against a pool of representative read-pileups from the original MRE clusters. In each iteration of the simulation, a randomly generated sequence with the same read-weighted base composition, and the same length and average coverage is generated for each significantly expressed MRE cluster. This pool of simulated CLIP-seq reads is then used to generate an expected count distribution for each MRE-motif with a non-zero observed count value. The total number of expected counts for iteration i is the cumulative number of reads present within the pool of simulated reads that harbor the MRE sequence of the seed-region formation (expected count c_i for miRNA j). The process was carried out one million times for each enumerated seed-region formation in turn in order to build a distribution of expected occurrences for the MRE-motif. The p-values for the enumerated seed-region formation were then calculated by fitting the expected count distribution of the variant with a negative binomial distribution (Supp. Fig. 2). Multiple test correction was performed using the Benjamini-Hochberg procedure and only those MRE-motifs with an FDR ≤ 0.05 were deemed to be significant and kept for further analysis. To enable a direct comparison between our work and those earlier efforts in which only *exonic* MRE clusters were considered and analyzed from the standpoint of miRNA targeting formations for miR-124[52], we repeated our Monte-Carlo simulations for the mouse brain samples considering only the set of *exonic* MRE clusters (instead of the full genome-wide set of MRE clusters). To this end, we first identified the MREs that are located within exonic regions and recomputed MRE significance. Then, we sub-selected and processed only statistically significant MREs (p-value ≤ 0.05). During the shuffling phase of the Monte Carlo simulation, the sub-selected exonic MREs could only be repositioned (shuffled) to other exonic regions within the mouse genome.

Selecting among 'competing' seed-region formations. Since our analysis transcends the standard model, on occasion we may find multiple seed-region formations competing for the same MRE. For example: miRNA X may match a given segment of an MRE using a variant containing multiple G:U wobbles in the seed region whereas miRNA Y may match the *exact same* segment of the *same* MRE using a variant that incorporates a bulge in the seed region. We resolve such conflicts with a multi-tiered approach. First, we filter the candidate seed-region formations using their associated False Discovery Rate (FDR): only variants with FDR ≤ 0.05 are considered significant. Second, we take into account the part beyond the seed of the miRNA that competes for a given MRE and examine how well and how extensively the full-length candidate miRNA base-pairs with the region that is adjacent and immediately upstream of the MRE at hand. To this end, we form full-length miRNA:target heteroduplexes using the sequence of each miRNA and a 25-nt stretch of the genome whose 3′ end extends one nucleotide past the 6-nt segment of the MRE-motif at hand using the Vienna package[77]. On the output of the co-folding we impose two additional constraints: first, we discard heteroduplexes whose Vienna-derived seed-region interactions do not match the sequence composition and architecture that are expected by the seed-region formation being considered; and, second, we discard heteroduplexes that contain instances of self-hybridization or comprise fewer than 12 base pairs. Results obtained from biological replicates of the same cellular phenotype were pooled together and duplicate entries removed.

1. Bartel, D. P. MicroRNAs: target recognition and regulatory functions. *Cell* **136**, 215–233 (2009).
2. Wightman, B., Ha, I. & Ruvkun, G. Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans. *Cell* **75**, 855–862 (1993).
3. Lee, R. C., Feinbaum, R. L. & Ambros, V. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* **75**, 843–854 (1993).
4. Reinhart, B. J. *et al.* The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. *Nature* **403**, 901–906 (2000).
5. Tay, Y., Zhang, J., Thomson, A. M., Lim, B. & Rigoutsos, I. MicroRNAs to Nanog, Oct4 and Sox2 coding regions modulate embryonic stem cell differentiation. *Nature* **455**, 1124–1128 (2008).
6. Poliseno, L. *et al.* A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. *Nature* **465**, 1033–1038 (2010).
7. Calin, G. A. & Croce, C. M. Chronic lymphocytic leukemia: interplay between noncoding RNAs and protein-coding genes. *Blood* **114**, 4761–4770 (2009).
8. Spizzo, R., Nicoloso, M. S., Croce, C. M. & Calin, G. A. SnapShot: MicroRNAs in Cancer. *Cell* **137**, 586–586 e581 (2009).
9. Rigoutsos, I. & Tsirigos, A. in *MicroRNAs in Development and Cancer* (ed Slack, F.) Ch. 10, 237–259 (Imperial College Press, 2010).
10. Ha, I., Wightman, B. & Ruvkun, G. A bulged lin-4/lin-14 RNA duplex is sufficient for Caenorhabditis elegans lin-14 temporal gradient formation. *Genes Dev* **10**, 3041–3050 (1996).
11. Vella, M. C., Choi, E. Y., Lin, S. Y., Reinert, K. & Slack, F. J. The C. elegans microRNA let-7 binds to imperfect let-7 complementary sites from the lin-41 3′UTR. *Genes Dev* **18**, 132–137 (2004).
12. Easow, G., Teleman, A. A. & Cohen, S. M. Isolation of microRNA targets by miRNP immunopurification. *RNA* **13**, 1198–1204 (2007).
13. Didiano, D. & Hobert, O. Perfect seed pairing is not a generally reliable predictor for miRNA-target interactions. *Nat Struct Mol Biol* **13**, 849–851 (2006).
14. Xia, Z. *et al.* Molecular dynamics simulations of Ago silencing complexes reveal a large repertoire of admissible 'seed-less' targets. *Sci Rep* **2**, 569 (2012).
15. Tay, Y. M. *et al.* MicroRNA-134 modulates the differentiation of mouse embryonic stem cells, where it causes post-transcriptional attenuation of Nanog and LRH1. *Stem Cells* **26**, 17–29 (2008).
16. Rigoutsos, I. New tricks for animal microRNAS: targeting of amino acid coding regions at conserved and nonconserved sites. *Cancer Res* **69**, 3245–3248 (2009).
17. Miranda, K. C. *et al.* A pattern-based method for the identification of MicroRNA binding sites and their corresponding heteroduplexes. *Cell* **126**, 1203–1217 (2006).
18. Selbach, M. *et al.* Widespread changes in protein synthesis induced by microRNAs. *Nature* **455**, 58–63 (2008).
19. Baek, D. *et al.* The impact of microRNAs on protein output. *Nature* **455**, 64–71 (2008).
20. Lal, A. *et al.* miR-24 Inhibits cell proliferation by targeting E2F2, MYC, and other cell-cycle genes via binding to "seedless" 3′UTR microRNA recognition elements. *Mol Cell* **35**, 610–625 (2009).
21. Thomas, M., Lieberman, J. & Lal, A. Desperately seeking microRNA targets. *Nat Struct Mol Biol* **17**, 1169–1174 (2010).
22. Cui, Y. H. *et al.* miR-503 represses CUG-binding protein 1 translation by recruiting CUGBP1 mRNA to processing bodies. *Mol Biol Cell* **23**, 151–162 (2012).
23. Chiang, K., Sung, T. L. & Rice, A. P. Regulation of cyclin T1 and HIV-1 Replication by microRNAs in resting CD4+ T lymphocytes. *J Virol* **86**, 3244–3252 (2012).
24. Sotillo, E. *et al.* Myc overexpression brings out unexpected antiapoptotic effects of miR-34a. *Oncogene* **30**, 2587–2594 (2011).
25. Liu, C. *et al.* The microRNA miR-34a inhibits prostate cancer stem cells and metastasis by directly repressing CD44. *Nat Med* **17**, 211–215 (2011).
26. Gao, J. S. *et al.* The Evi1, microRNA-143, K-Ras axis in colon cancer. *FEBS Lett* **585**, 693–699 (2011).
27. Takagi, S. *et al.* MicroRNAs regulate human hepatocyte nuclear factor 4alpha, modulating the expression of metabolic enzymes and cell cycle. *J Biol Chem* **285**, 4415–4422 (2010).
28. Abdelmohsen, K., Srikantan, S., Kuwano, Y. & Gorospe, M. miR-519 reduces cell proliferation by lowering RNA-binding protein HuR levels. *Proc Natl Acad Sci U S A* **105**, 20297–20302 (2008).
29. Fasanaro, P. *et al.* ROD1 is a seedless target gene of hypoxia-induced miR-210. *PLoS One* **7**, e44651 (2012).
30. Adilakshmi, T., Sudol, I. & Tapinos, N. Combinatorial action of miRNAs regulates transcriptional and post-transcriptional gene silencing following in vivo PNS injury. *PLoS One* **7**, e39674 (2012).
31. Surdziel, E. *et al.* Enforced expression of miR-125b affects myelopoiesis by targeting multiple signaling pathways. *Blood* **117**, 4338–4348 (2011).
32. Duursma, A. M., Kedde, M., Schrier, M., le Sage, C. & Agami, R. miR-148 targets human DNMT3b protein coding region. *RNA* **14**, 872–877 (2008).
33. Lal, A. *et al.* p16(INK4a) translation suppressed by miR-24. *PLoS One* **3**, e1864 (2008).
34. Stark, A. *et al.* Discovery of functional elements in 12 Drosophila genomes using evolutionary signatures. *Nature* **450**, 219–232 (2007).
35. Chi, S. W., Zang, J. B., Mele, A. & Darnell, R. B. Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. *Nature* **460**, 479–486 (2009).
36. Hafner, M. *et al.* Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* **141**, 129–141 (2010).
37. Konig, J. *et al.* iCLIP–transcriptome-wide mapping of protein-RNA interactions with individual nucleotide resolution. *J Vis Exp* (2011).
38. Helwak, A., Kudla, G., Dudnakova, T. & Tollervey, D. Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding. *Cell* **153**, 654–665 (2013).
39. Majoros, W. H. *et al.* MicroRNA target site identification by integrating sequence and binding information. *Nat Methods* **10**, 630–633 (2013).
40. Khorshid, M., Hausser, J., Zavolan, M. & van Nimwegen, E. A biophysical miRNA-mRNA interaction model infers canonical and noncanonical targets. *Nat Methods* **10**, 253–255 (2013).
41. Hafner, M., Lianoglou, S., Tuschl, T. & Betel, D. Genome-wide identification of miRNA targets by PAR-CLIP. *Methods* **58**, 94–105 (2012).
42. Schug, J. *et al.* Dynamic recruitment of microRNAs to their mRNA targets in the regenerating liver. *BMC Genomics* **14**, 264 (2013).
43. Chou, C. H. *et al.* A computational approach for identifying microRNA-target interactions using high-throughput CLIP and PAR-CLIP sequencing. *BMC Genomics* **14 Suppl 1**, S2 (2013).
44. Corcoran, D. L. *et al.* PARalyzer: definition of RNA binding sites from PAR-CLIP short-read sequence data. *Genome Biol* **12**, R79 (2011).

45. Erhard, F., Dolken, L., Jaskiewicz, L. & Zimmer, R. PARma: identification of microRNA target sites in AGO-PAR-CLIP data. *Genome Biol* **14**, R79 (2013).
46. Murigneux, V., Sauliere, J., Roest Crollius, H. & Le Hir, H. Transcriptome-wide identification of RNA binding sites by CLIP-seq. *Methods* **63**, 32–40 (2013).
47. Li, J. H., Liu, S., Zhou, H., Qu, L. H. & Yang, J. H. starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. *Nucleic Acids Res* **42**, D92–97 (2014).
48. Liu, C. *et al.* CLIP-based prediction of mammalian microRNA binding sites. *Nucleic Acids Res* **41**, e138 (2013).
49. Wang, W. X. *et al.* The expression of microRNA miR-107 decreases early in Alzheimer's disease and may accelerate disease progression through regulation of beta-site amyloid precursor protein-cleaving enzyme 1. *J Neurosci* **28**, 1213–1223 (2008).
50. Kishore, S. *et al.* A quantitative analysis of CLIP methods for identifying binding sites of RNA-binding proteins. *Nat Methods* **8**, 559–564 (2011).
51. Pillai, M. M. *et al.* HITS-CLIP reveals key regulators of nuclear receptor signaling in breast cancer. *Breast Cancer Res Treat* **146**, 85–97 (2014).
52. Chi, S. W., Hannon, G. J. & Darnell, R. B. An alternative mode of microRNA target recognition. *Nat Struct Mol Biol* **19**, 321–327 (2012).
53. Loeb, G. B. *et al.* Transcriptome-wide miR-155 binding map reveals widespread noncanonical microRNA targeting. *Mol Cell* **48**, 760–770 (2012).
54. Huang da, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**, 44–57 (2009).
55. Dennis, G., Jr. *et al.* DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol* **4**, P3 (2003).
56. Li, T. *et al.* miR-155 regulates the proliferation and cell cycle of colorectal carcinoma cells by targeting E2F2. *Biotechnol Lett* (2014).
57. Seddiki, N., Brezar, V., Ruffin, N., Levy, Y. & Swaminathan, S. Role of miR-155 in the regulation of lymphocyte immune function and disease. *Immunology* **142**, 32–38 (2014).
58. Hsu, S. D. *et al.* miRTarBase: a database curates experimentally validated microRNA-target interactions. *Nucleic Acids Res* **39**, D163–169 (2011).
59. Esquela-Kerscher, A. & Slack, F. J. Oncomirs - microRNAs with a role in cancer. *Nat Rev Cancer* **6**, 259–269 (2006).
60. Mendell, J. T. miRiad roles for the miR-17-92 cluster in development and disease. *Cell* **133**, 217–222 (2008).
61. Mogilyansky, E. & Rigoutsos, I. The miR-17/92 cluster: a comprehensive update on its genomics, genetics, functions and increasingly important and numerous roles in health and disease. *Cell Death Differ* **20**, 1603–1614 (2013).
62. Ziu, M. *et al.* Spatial and temporal expression levels of specific microRNAs in a spinal cord injury mouse model and their relationship to the duration of compression. *Spine J* **14**, 353–360 (2014).
63. Jalali, S., Bhartiya, D., Lalwani, M. K., Sivasubbu, S. & Scaria, V. Systematic transcriptome wide analysis of lncRNA-miRNA interactions. *PLoS One* **8**, e53823 (2013).
64. Paraskevopoulou, M. D. *et al.* DIANA-LncBase: experimentally verified and computationally predicted microRNA targets on long non-coding RNAs. *Nucleic Acids Res* **41**, D239–245 (2013).
65. Ebert, M. S. & Sharp, P. A. Emerging roles for natural microRNA sponges. *Curr Biol* **20**, R858–861 (2010).
66. Ebert, M. S. & Sharp, P. A. MicroRNA sponges: progress and possibilities. *RNA* **16**, 2043–2050 (2010).
67. Salmena, L., Poliseno, L., Tay, Y., Kats, L. & Pandolfi, P. P. A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? *Cell* **146**, 353–358 (2011).
68. Vourekas, A. *et al.* Mili and Miwi target RNA repertoire reveals piRNA biogenesis and function of Miwi in spermiogenesis. *Nat Struct Mol Biol* **19**, 773–781 (2012).
69. Kiriakidou, M. *et al.* A combined computational-experimental approach predicts human microRNA targets. *Genes Dev* **18**, 1165–1178 (2004).
70. Leung, A. K. *et al.* Genome-wide identification of Ago2 binding sites from mouse embryonic stem cells with and without mature microRNAs. *Nat Struct Mol Biol* **18**, 237–244 (2011).
71. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. journal* **17**, 10–12 (2011).
72. David, M., Dzamba, M., Lister, D., Ilie, L. & Brudno, M. SHRiMP2: sensitive yet practical SHort Read Mapping. *Bioinformatics* **27**, 1011–1012 (2011).
73. Kozomara, A. & Griffiths-Jones, S. miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res* **39**, D152–157 (2011).
74. Xu, J. & Zhang, Y. A generalized linear model for peak calling in ChIP-Seq data. *J Comput Biol* **19**, 826–838 (2012).
75. Haecker, I. *et al.* Ago HITS-CLIP expands understanding of Kaposi's sarcoma-associated herpesvirus miRNA function in primary effusion lymphomas. *PLoS Pathog* **8**, e1002884 (2012).
76. Zisoulis, D. G. *et al.* Comprehensive discovery of endogenous Argonaute binding sites in Caenorhabditis elegans. *Nat Struct Mol Biol* **17**, 173–179 (2010).
77. Bernhart, S. H. *et al.* Partition function and base pairing probabilities of RNA heterodimers. *Algorithms Mol Biol* **1**, 3 (2006).

Acknowledgments

The authors wish to thank Eleftheria Hatzimichael, for helpful feedback and stimulating discussions during the length of this project. This research was supported in part by the William M. Keck Foundation (IR), the Hirshberg Foundation for Pancreatic Cancer Research (IR and JB), NIH-NIAID (2U19AI056363-06/2030984 to IR), by institutional funds, and in part by a grant to IR from the Pennsylvania Department of Health which specifically disclaims responsibility for any analyses, interpretations or conclusions. The research was also supported by the National Cancer Institute of the National Institutes of Health under Award Number P30CA056036. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author contributions

IR spearheaded and supervised the project. P.C. and I.R. conceived and designed the experiments. P.C., P.L. and I.R. wrote the analysis software. K.Q. performed laboratory experiments. P.C., I.R., P.L. and K.Q. carried out the analyses. J.B. assisted with sample collection and HITS-CLIP, and reviewed the data. P.C., I.R., E.L. and P.L. wrote the manuscript. All authors reviewed and contributed edits to the manuscript.

Additional information

Data Availability The sequence data for the hTERT-HPNE and MIA PaCa-2 HITS-CLIP are available on GEO under accession # SRP034075.

Supplementary information accompanies this paper at http://www.nature.com/scientificreports

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Clark, P.M. *et al.* Argonaute CLIP-Seq reveals miRNA targetome diversity across tissue types. *Sci. Rep.* **4**, 5947; DOI:10.1038/srep05947 (2014).