Integrated analysis of directly captured microRNA targets reveals the impact of microRNAs on mammalian transcriptome

GLEN A. BJERKE and RUI YI
Department of Molecular, Cellular and Developmental Biology, University of Colorado Boulder, Boulder, Colorado 80309, USA

ABSTRACT
MicroRNA (miRNA)-mediated regulation is widespread, relatively mild but functionally important. It remains challenging to unequivocally identify miRNA targeted RNAs at a genomic scale and determine how changes in miRNA levels affect the transcriptome. Here, we captured individual miRNAs and their targeted RNA sites in wild-type, miR-200 family knock-out and induced epithelial cells. We detected 1797 miRNAs interacting with 13,830 transcripts at 616,127 sites by sequencing 1,230,019 unique miRNA:RNA chimeras. Although mRNA sites that are bound by miRNAs and contain matches to seed sequences confer the strongest regulation, ∼40%–60% of miRNA bound regions do not contain seed matches. Different miRNAs have different preferences to seed matches and 3′ end base-pairing. For individual miRNAs, the effectiveness of mRNA regulation is highly correlated with the number of captured miRNA:mRNA chimeras. Notably, elevated miR-200 expression robustly represses existing targets with little impact on newly recognized targets. Global analysis of directly captured mRNA targets reveals pathways that are involved in cancer and cell adhesion and signaling pathways that are highly regulated by many different miRNAs in epithelial cells. Comparison between experimentally captured and TargetScan predicted targets indicates that our approach is more effective in identifying bona fide targets by reducing false positive and negative predictions. This study reveals the global binding landscape and impact of miRNAs on the mammalian transcriptome.

Keywords: miRNA binding; direct capture of miRNA targets; transcriptome

INTRODUCTION
MicroRNAs (miRNAs) are small noncoding RNAs that repress gene expression through sequence-specific mRNA target binding, resulting in mRNA degradation or repression of translation (Jonas and Izaurralde 2015). The majority of mammalian genes are thought to be regulated by miRNAs, suggesting that miRNA-mediated gene expression regulation is widespread and functionally important (Ambros 2004; Bartel 2018). However, understanding the function of an individual miRNA or miRNA family has been hampered by the fact that a single miRNA may regulate many genes and a single gene can be regulated by many different miRNAs (Baumjohann and Ansel 2013). In addition, a single miRNA:mRNA pairing usually results in mild regulation of gene expression, indicating that multiple interactions are necessary to impart strong repression, either against one gene or multiple components within a pathway (Baumjohann and Ansel 2013). This also indicates that to comprehensively understand the effect of a miRNA or miRNA family, the regulation of the entire transcriptome has to be quantitatively measured as opposed to one or a few targets of interest in a cell type-specific manner (Gurtan and Sharp 2013; Bleazard et al. 2015). Furthermore, when the expression levels of miRNAs elevate in conditions such as normal development, stem cell self-renewal and differentiation, pathogenesis or through experimental manipulation, it is unclear whether the increased miRNA expression causes stronger binding and repression to canonical targets or if this results in de novo miRNA:mRNA interactions with new targets. Distinguishing these two possibilities is particularly important to understand mechanisms that underlie robust

© 2020 Bjerke and Yi This article is distributed exclusively by the RNA Society for the first 12 months after the full-issue publication date (see http://majournal.cshlp.org/site/misc/temps.xhtml). After 12 months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at http://creativecommons.org/licenses/by-nc/4.0/.
phenotypes usually caused by increased miRNA expression in many gain-of-function studies.

miRNA:RNA interaction is mediated by partial base-pairing between miRNA and RNA sequences. During miRNA biogenesis in mammals, a single-stranded mature miRNA is loaded onto one of the four Argonaute (Ago) proteins, usually determined by the protein abundance (Wang et al. 2012), to form the RNA induced silencing complex (RISC) (Thomson et al. 2011). Structural, biochemical, and computational studies have all demonstrated that the sequences at the 5’ end of miRNAs, often termed the seed region, are most critical to miRNA:mRNA interactions (Bartel 2009) although miRNA:mRNA interactions via the 3’ regions of miRNAs also likely play a role (Moore et al. 2015). Over the past 15 yr, extensive efforts have been dedicated to develop computational tools, which are generally based on identifying miRNA regions with a seed match, to identify miRNA targets. Prominent algorithms such as TargetScan have become a major resource for miRNA target prediction (Bartel 2018). However, computational predictions generally cannot account for variable cellular contexts or detect targets with nonperfectly matched seed sequences and they also lack the ability to distinguish functionally important target genes or pathways, usually due to a large number of predicted targets (Pinzón et al. 2017). Given the difficulty of predicting targets, many experimental methods for determining targets have been developed. For example, the HITS-CLIP method has been established to identify miRNA target sites that are crosslinked by UV radiation and associated with Ago proteins (Chi et al. 2009). Because miRNA recognized target sites are bound and protected by Ago proteins, they are subsequently recovered from Ago immunoprecipitation and deeply sequenced for identification of miRNA targets. PAR-CLIP was further developed to improve the crosslink efficiency of HITS-CLIP by incorporating 4-thiouridine into the RNA in cultured cells (Hafner et al. 2010). Both methods, however, do not identify miRNAs that mediate the recognition of mRNA target sites and often detect large mRNA regions that are associated with Ago proteins. As a result, extensive bioinformatic analysis, often relying on the presence of miRNA seed matches (6mer to 8mer sequences) within mRNA sequences, has to be used to identify miRNA binding sites and assign them to specific miRNAs (Chi et al. 2009; Hafner et al. 2010). Because of the prevalence of 6mer to 8mer sequences in mammalian transcriptome, such an identification often generates false positives and also cannot assign mRNA regions without a match to canonical miRNA seed regions. To overcome these difficulties, the CLASH technique was reported to identify miRNA targets by sequencing ligated miRNA:RNA chimeras, which allows the identification of single miRNA and its associated mRNA site regardless of the presence of seed matches (Kudla et al. 2011; Helwak et al. 2013; Broughton et al. 2016). The low frequency of miRNA:mRNA chimeric reads in the original CLASH study was then improved upon by the CLEAR-CLIP method that adds an additional ligation step to enhance miRNA:RNA ligation (Moore et al. 2015). This technique allows identification of bona fide miRNA targets through direct capture and sequencing of miRNA:RNA pairs that were ligated while still in the RISC and removes the imperfect bioinformatic prediction that assigns a miRNA to an mRNA site as required by HITS-CLIP (Moore et al. 2015). However, it remains unclear whether CLEAR-CLIP can distinguish strong versus weak miRNA target sites and how CLEAR-CLIP identified miRNA:mRNA interactions reflect miRNA-mediated mRNA regulation.

In this study, we utilize improved experimental conditions that allow more efficient and unbiased RNA ligation (Zhang et al. 2013) to enhance the ability of CLEAR-CLIP to capture miRNA:RNA chimeras. By using randomized adapters and PEG-8000 during the ligation steps, we further improve the quantitative performance of CLEAR-CLIP. To explore the impact of miRNAs on the transcriptome of mammalian cells, we apply this method to keratinocytes derived from mouse skin with a focus on the miR-200 family. The five members of miR-200s differ by a single nucleotide within the seed region and also have variable sequences at the 3’ end (Supplemental Fig. S1A). Thus, a highly accurate target identification method is required to distinguish how each member of the family recognizes and regulates their targets. In addition to wild-type cells, we use primary epithelial cells isolated from miR-200 double knockout mice, in which all five members of the miR-200 family are deleted, and miR-200 inducible mice, in which three members of the family are induced (Hoefer et al. 2018). These data are also combined with RNA-seq of miR-200 induced cells, allowing determination of which miRNA:mRNA interactions are functional in repressing mRNA levels. We further validate our method with miR-205, one of the most highly expressed miRNAs in epithelial stem cells (Wang et al. 2013). Additionally, we show that CLEAR-CLIP can be used to identify unique and common targets of miRNA isoforms that differ by a single nucleotide at their 5’ end sequences. The lessons learned from these individual miRNAs are then applied to the miRNA pathway globally, generating new knowledge of miRNA-regulated networks and how they control pathways involved in cancer, cell adhesion and signaling in epithelial cells of the skin. Together, this work establishes an experimental framework to accurately capture all miRNA targets in a miRNA- and target site-specific manner. By comprehensively mapping all miRNA:RNA interactions genomewide, this approach should facilitate mechanistic studies of miRNA-mediated gene expression regulation in development, homeostasis, stress response, and disease.
RESULTS

CLEAR-CLIP identifies miRNA site-specific interactions at a genomic scale

We previously generated mouse skin specific models for double knockout of the two miR-200 clusters (Krt14-Cre/miR-200c:141′/‘miR-200b:200a:429′/‘, hereafter referred to as DKO) and transgenic miR-200 inducible overexpression (Krt14-rtTA/pTRE2-miR-200b:200a:429, hereafter referred to as miR-200 Tg) (Hoefert et al. 2018). We performed CLEAR-CLIP on mouse keratinocytes from nine control samples, six miR-200 DKO samples and three miR-200 Tg samples, allowing us to validate loss or gain of CLEAR-CLIP signals (samples detailed in Supplemental Table S1). We performed an optimized version of CLEAR-CLIP to enhance quantitativeness and sensitivity of detection (see Fig. 1A and Materials and Methods). To reduce bias in the ligation steps and enhance ligation efficiency, we used a 5′ linker with a random NNNN and a 3′ linker with a NN at the ligated end of each adaptor and performed the ligation steps in the presence of PEG8000. These steps were effective at reducing bias and improving ligation efficiency for small RNA ligation and sequencing (Zhang et al. 2013). CLEAR-CLIP reads were barcoded using the 5′ NNNN to distinguish unique events, allowing removal of PCR duplicates and in total we sequenced 1,230,019 unique mRNA:RNA chimeras from 18 libraries. To assure robust target detection, we required a mRNA region to be ligated to the same miRNA in at least two libraries, which we will hereafter refer to as “high-confidence” sites.

miRNAs can affect both mRNA degradation and translational repression; however, the measurement of miRNA levels has been shown to be an effective measurement of miRNA regulation in mammalian cells (Baek et al. 2008; Guo et al. 2010). We combined the CLEAR-CLIP results with RNA-seq data from miR-200 Tg (Supplemental Fig. S1B) and miR-205 induced mouse keratinocytes (Krt14-rtTA/pTRE2-miR-205, hereafter referred to as miR-205 Tg) (Supplemental Fig. S1C) to characterize mRNA targeting and expression by the miR-200 family, miR-205 and the entire miRNA pathway.

For the miR-200 family, we identified 2352 high-confidence miR-200 sites in 3′ UTRs of mRNAs, corresponding to 1486 unique genes. We first compared the sensitivity of detecting miRNA targeted sites between mRNA only reads that are typically obtained in HITS-CLIP and miRNA:mRNA chimeric reads that are obtained by CLEAR-CLIP. When we examined the relative read density of mRNA only fragments within these miR-200 high-confidence sites as defined by miRNA:mRNA chimeras, we observed little to no difference between control, miR-200 Tg and DKO samples (Fig. 1B). In contrast, when we used miR-200:mRNA chimeric reads to specifically examine miR-200-mediated targeting events, we observed an increase in miR-200 Tg samples (~1.25-fold) and an almost total loss in the DKO samples as expected (Fig. 1C). These results indicate that examining mRNA only reads as in HITS-CLIP is not sensitive enough to detect changes in mRNA levels, but the change can be detected using chimeric CLEAR-CLIP reads. As a control, we also calculated read density within high-confidence sites of miR-205, a highly expressed but unrelated miRNA, in control, miR-200 Tg and DKO samples. When using mRNA only reads, we did not observe any change in mRNA coverage for these miR-205 sites between control and miR-200 Tg and a slight increase (~1.25-fold) in the miR-200 DKO (Fig. 1D). When miR-205:mRNA chimeric reads were examined, there was again no change between control and miR-200 Tg but a larger increase in the miR-200 DKO sample (Fig. 1E). These data indicate higher sensitivity and quantitative performance of CLEAR-CLIP relative to HITS-CLIP. Furthermore, these results suggest that in the absence of miR-200s there is an increase in targeting by other miRNAs.

We also compared our CLEAR-CLIP data to HITS-CLIP previously published by our laboratory (Riemondy et al. 2015) and found that for both miR-200s and miR-205, CLEAR-CLIP identified more miRNA targets that resulted in better gene repression (Supplemental Fig. S1D,E). Together, these data indicate the high sensitivity and specificity of CLEAR-CLIP for identifying miRNA-interacting sites in comparison to HITS-CLIP.

We next examined the genomic distribution of miR-200 high-confidence sites. When requiring targeted sites to be found in two or more libraries, ~45% of sites annotated to 3′ UTRs, 35% to introns and small percentages to 5′ UTRs, CDS, miRNAs, noncoding RNAs (ncRNAs) and intergenic regions (Fig. 1F). However, ~70% of sites annotated to 3′ UTRs and ~10% annotated to introns when requiring sites to be found in 5+ or 8+ libraries (Fig. 1F). These results indicate that even though chimeras to intronic regions are seen with some frequency, reproducible sites are usually found in 3′ UTRs. Additionally, to determine whether miR-200 sites from different regions of the genome were functional, we selected genes that had only one miR-200 high-confidence site and analyzed their expression when miR-200s were induced. We found 3′ UTR sites to be highly effective, while CDS sites had a slight but statistically significant effect and other regions including 5′ UTR and intron were not effective for gene repression (Supplemental Fig. S1F). miR-200 high-confidence sites within 3′ UTRs, as opposed to other portions of the genome, contained the highest percentage of motifs (Supplemental Fig. S1G). Notably, 5′ UTRs contained the lowest percentage of motif matches (Supplemental Fig. S1G). We subsequently focused our studies on 3′ UTR sites.

Because miR-200 miRNAs from the two sub-families differ by one nucleotide at the fourth position in their seed
sequences, we examined how many individual miR-200 family members recognize the same site. The majority of miR-200 sites were targeted by one (35% of sites) or two (42.5% of sites) family members, but some sites were targeted by three (17.5%), four (3.8%), or all five (1.2%) family members (Fig. 1G). The majority of the sites (80.7%) had reads from the same seed family, while the rest (19.3%) had reads from both miR-200 seed sub-families (Supplemental Fig. S1H).
the opposite seed. Overall, approximately 60% of high-confidence sites for all miR-200 family members contained their cognate seed and ~30% did not contain either seed (Fig. 1H). A small percentage of sites contained either both seeds or the opposite seed, indicating that miR-200 family members are much more likely to bind to a site with the cognate seed even though their seed sequences only differ by one nucleotide. As shown in Supplemental Figure S1I, we also observed individual members of the miR-200 family binding to similar percentages of motifs although miR-200a binds to more 6mers than other family members.

The specificity of each family member was further examined using RNAhybrid (Rehmsmeier et al. 2004) to calculate the predicted binding energy for all five family members against sites that were found to be dominated by an individual family member (requiring >50% of reads within a high-confidence area to come from one family member). Most sites that had a majority of reads from one family member correspondingly had the lowest binding energy for that member, and seed families also had lower binding energy within their family (Fig. 1I). Interestingly, miR-200c and miR-429 had lower average binding energies to miR-200b sites than miR-200b itself, likely due to the lower binding energy of miR-200c and miR-429 to their perfect reverse complementary target than miR-200b (~43.3 kcal/mol and ~40.3 kcal/mol, respectively vs. ~38.5 kcal/mol for miR-200b).

To probe the genes regulated by this five-member family, we performed pathway analysis on their high-confidence targets. Using high-confidence sites with a seed (7mer or better) in 3’ UTRs we compiled a gene list for each miR-200 family member. Gene Ontology (GO) was then performed on these gene lists examining for enrichment of KEGG pathways using Enrichr (Kuleshov et al. 2016). Focal adhesion, Ras signaling and PI3K-Akt signaling are the best miR-200 high-confidence site with or without a miR-200 seed (6mer or better) in 3’ UTRs, we performed pathway analysis on their high-confidence targets. Using high-confidence sites with a seed (7mer or better) in 3’ UTRs we compiled a gene list for each miR-200 family member. Gene Ontology (GO) was then performed on these gene lists examining for enrichment of KEGG pathways using Enrichr (Kuleshov et al. 2016). Focal adhesion, Ras signaling and PI3K-Akt signaling pathways were found to be regulated by both seed families, supporting a coordinated targeting mechanism by the miR-200 family (Hoefert et al. 2018).

**CLEAR-CLIP identifies functional miRNA targeting sites**

miRNA levels often dramatically change during homeostasis such as cell fate specification, during stress responses such as wound healing as well as under pathological conditions such as tumorigenesis (Mendell and Olson 2012; Lin and Gregory 2015). Furthermore, many studies have relied on overexpression to examine miRNA functions. However, it is unknown whether elevated miRNA expression preferentially represses existing targets or inhibits new targets (Thomson et al. 2011). To address this issue, we examined whether overexpression of the miR-200b/a/429 cluster above physiological levels (approximately eight- to 15-fold) results in off-target effects, causing miRNAs to target new genes not seen at physiological miRNA levels. To accomplish this, normalized read numbers in miR-200 high-confidence sites were calculated for control versus induced samples. There was a shift toward higher relative reads numbers in induced samples, as expected, and we also found a relatively small number of sites that were found only in the control or only in the induced (Fig. 2A). Upon induction of the miR-200b cluster, we observed that genes identified in both control and induced samples (1091 genes) were better repressed at the mRNA level than genes containing sites seen only in control (98 genes) or miR-200 induced samples (252 genes) (Fig. 2B). The stronger repression of shared sites was further supported by the observation that these sites had more CLEAR-CLIP reads with higher miR-200 expression in induced samples (Fig. 2C). We also observed that areas found in both control and induced samples were more likely to contain seed motifs ( ~75% vs. ~50% for control only or inducible only) (Supplemental Fig. S2A) and genes found in both control and induced samples tended to be slightly higher expressed (Supplemental Fig. S2B). These data indicate that elevated expression of miRNAs primarily results in stronger targeting of canonical sites.

Because ~30% of miR-200 CLEAR-CLIP high-confidence areas did not contain a seed match (Fig. 1H), we next examined the effectiveness of CLEAR-CLIP identified target sites in mediating gene repression. Genes with a miR-200 high-confidence site with or without a miR-200 seed (6mer or better) were both significantly repressed by induction of miR-200s. However, genes with the seed motif were repressed significantly better at the mRNA level than genes without a motif (Fig. 2D). Interestingly, genes containing CLEAR-CLIP identified target sites of other miRNAs were slightly derepressed upon induction of miR-200s, indicating competition for the availability of the RISC by induced miR-200s (Saito and Sætrom 2012). Additionally, genes that were derepressed (>0.05 log2 fold change) tended to have lower expression level (Supplemental Fig. S2C) and more total CLEAR-CLIP reads (Supplemental Fig. S2D) as compared to genes that were not derepressed.

To further analyze the impact of different types of seed matches, we classified miR-200 targets into categories by the best miRNA motif they contained (8mer > 7merM8 > 7merA1 > 6mer). We then performed enrichment of KEGG pathways using Enrichr (Kuleshov et al. 2016). Focal adhesion, Ras signaling and PI3K-Akt signaling pathways were found to be better repressed with the stronger match conferring stronger inhibition, consistent with previous reports (Bartel 2018). Statistically significant repression was also seen for genes with a 6mer match or without any seed match, but the repression for both groups of genes was much weaker (Fig. 2E).
Furthermore, genes with multiple miR-200 CLEAR-CLIP sites with a 6mer or without any seed match were not repressed more than genes with a single 6mer or no motif, reflecting the general ineffectiveness of 6mer and nonmotif miRNA targeting (Fig. 2F). These data suggest that, at least for miR-200s, a 7mer or 8mer match in a CLEAR-CLIP identified site is critical for effective repression at the mRNA level. In addition to these seed match-containing sites, we noted that many miR-200 bound regions without seed matches were still reproducibly seen in multiple control and induced libraries and absent from DKO libraries (Supplemental Fig. S2E). Because these sites did not usually contribute to the repression of miRNAs (Fig. 2E,F), they possibly represent a target scanning activity by the RISC.

Similarly, the effectiveness of seed matches was examined in miR-205 high-confidence sites. Compared to miR-200s, miR-205 sites showed fewer canonical 7mer or 8mer motifs and more sites that did not have a seed match. Therefore, we performed de novo motif analysis using...
HOMER (Heinz et al. 2010) on miR-205 sites that did not have a canonical seed match. This analysis found a 5mer match (nucleotides 3–7 of the miRNA) that was prevalent in 22% of the nonseed matching sites (Supplemental Fig. S3A). When we combined miR-205 CLEAR-CLIP identified sites with miR-205 Tg RNA-seq, we found the dependence on miR-205 seed matches less apparent than for miR-200s and only observed significant repression of genes that contained an 8mer or 7mer match (Fig. 2G). These data suggest that the effectiveness of a seed match may vary among different miRNAs.

We next examined whether miR-200s could target non-coding RNAs (ncRNAs) that are annotated in a database of murine ncRNAs (Fang et al. 2018). We found 1292 miR-200 high-confidence sites in 806 unique ncRNA genes. Among these, 597 ncRNA genes were detected in our poly(A) selected RNA-seq data. Notably, ncRNAs with miR-200 high-confidence sites containing an 8mer or 7mer were significantly repressed at the RNA level (Fig. 2H), similar to coding genes. Examples of miR-200s interacting with ncRNAs are shown in Supplemental Figure S3B, suggesting that miR-200s may also play a role in repressing ncRNAs. Since effective miRNA sites are usually found in 3’ UTRs of miRNAs, we next examined whether miR-200s were targeting a particular region on ncRNAs. However, no positional bias of miR-200 target sites along the length of ncRNAs was observed for either effective or ineffective sites (Fig. 2I). Because this result could be confounded by the length of the ncRNA, we further examined the distance of miR-200 sites to the 3’ end of ncRNAs and did not observe a bias of effective sites toward the 3’ end (Supplemental Fig. S3C).

Different miRNAs have a different degree of reliance on the seed match

Because the miR-200 family harbors two distinct seed sequences and these five miRNAs are co-expressed in epithelial cells, we assessed whether repression of miR-200 high-confidence targets was affected by containing one or both seed types. Genes that contained both a miR-200a and miR-200b type seed were repressed better upon miR-200 induction than genes that contained either one miR-200a type or miR-200b type seed (Fig. 3A). Genes with both seed types were repressed similarly to genes with two miR-200b type seeds, but significantly better than genes with two miR-200a type seeds. This is likely due to the fact that we induced expression of the miR-200b cluster that expresses two miR-200b type miRNAs and only one miR-200a type miRNA. We next examined whether genes that are targeted by more members of the miR-200 family are better repressed by induction of miR-200s in general. Indeed, the more miR-200 family members that were found to be associated with a gene by CLEAR-CLIP, the better repression upon miR-200 induction (Fig. 3B). These data suggest that CLEAR-CLIP can quantitatively measure the strength of miRNA-mediated repression.

To further characterize binding by miR-200s, high-confidence sites for each miR-200 member were generated individually and RNAhybrid was used to calculate the best binding site within each area. This information was then used to calculate how often each nucleotide was paired to its miRNA target. As shown in Figure 3C, miR-200s have a strong preference for a seed match. Interestingly, a large percentage of miR-200 areas also used nucleotides 12–14 of the miRNA for binding their targets. The binding fraction by nucleotide was also calculated for miR-205 (Fig. 3D). By comparison, miR-205 depends more heavily on nucleotides 3–7 of its seed, and utilizes more 3’ end binding (nucleotides 15–18) than miR-200s. These data suggest that 3’ end binding is variable among different miRNAs.

Next, the predicted binding from RNAhybrid was used to classify different binding subcategories by k-means clustering all binding subsets of the miR-200 family members and then examining their ability to repress gene expression. As shown in Figure 3E, miR-200s clustered into four mostly seed containing groups with different modes of 3’ end binding (groups 1, 3, 4, 5) and one group that lacked seed matches (group 2). Examining gene expression upon the induction of miR-200s, it was evident that the four groups with a seed match resulted in similar repression, whereas group 2, which lacked a seed match, resulted in less repression (Fig. 3F). We also performed the same analysis for miR-205 high-confidence targets (Fig. 3G) and again observed more variability in miR-205 binding using its seed region. Assaying the functionality of these groups using RNA-seq revealed groups 1, 2, 3, and 5, which contain both seed and 3’ end matches, were significantly repressed. Group 4 genes, which did not contain seed matches, were also significantly repressed, but less so (Fig. 3H). These data show that miR-200s are highly dependent on a seed match, whereas miR-205 is more dependent on both seed and 3’ end binding.

Quantitative analysis of CLEAR-CLIP identified miRNA targets

We next tested whether our optimized CLEAR-CLIP can predict the strength of miR-200-dependent regulation based on the number of captured miRNA:mRNA chimeric reads. We first examined the correlation between the number of miR-200 CLEAR-CLIP reads per gene and their repression upon miR-200 induction. Genes with increasing numbers of CLEAR-CLIP reads were more repressed at the mRNA level upon induction with miR-200s (Fig. 4A). Additionally, when miR-200 chimeric reads account for 20% or more of total CLEAR-CLIP reads for a gene, they also confer stronger repression by miR-200 induction.
FIGURE 3. miR-200s are highly dependent on seed matches for target repression. (A) Log2 fold change in mRNA expression upon induction of the miR-200b cluster is shown for genes that contain one miR-200a type seed, one miR-200b type seed, both seed types, or two or more of each seed type. (B) Log2 fold change in mRNA expression upon induction of the miR-200b cluster is shown for genes that contain two or more reads from one, two, three, four, or five miR-200s as compared to non-miR-200 peak genes. (C) The predicted fraction bound is displayed for each nucleotide along the length of each miR-200 for all of that member’s high-confidence areas. (D) The predicted fraction bound is displayed for each nucleotide along the length of miR-205. (E) High-confidence miR-200 areas with the majority of reads for that area from a single family member were hybridized using RNAhybrid and individual nucleotides were predicted to be bound or not bound, with the seed being enforced if one was present in the area. All miR-200 family members were then pooled, and the predicted nucleotide hybridization was clustered in five groups using k-means clustering. This hybridization was then graphed as a heat map with black denoting the nucleotide is bound and white meaning not bound. (F) Log2 fold change in mRNA expression upon induction of the miR-200b cluster is shown for the five k-means clusters from E, as compared to non-miR-200 peak genes. The indicated P-value for each is compared to non-miR-200 peak genes. (G) High-confidence miR-205 areas were hybridized against miR-205 using RNAhybrid, with the seed being enforced if one was present within the area. Predicted binding was then clustered into five k-means clusters and graphed as a heat map with black denoting the nucleotide is bound and white meaning not bound. The indicated P-value for each is compared to non-miR-205 peak genes. (H) Log2 fold change in gene expression upon induction of miR-205 is shown for the five k-means clusters from G, as compared to non-miR-205 peak genes. For all CDF and violin plots the number of genes is shown in parentheses and P-values were calculated using the Kolmogorov–Smirnov test.
We further binned miR-200 targets into three groups based on reads per gene (one to six, seven to 30, or 31+ reads) and further divided each group into two categories: low percent miR-200 targeting, where miR-200s account for <20% of the total CLEAR-CLIP reads for the gene, or high percent miR-200 targeting, where miR-200s account for >20% of the total CLEAR-CLIP reads for the gene. Interestingly, high percentage miR-200 targeting was correlated with better mRNA repression for each group (Fig. 4C).

Similarly, for miR-205, more miR-205 CLEAR-CLIP reads per gene was also correlated with stronger repression. Genes with three to nine or 10+ reads were significantly repressed at the mRNA level by miR-205 induction, in contrast to genes with only one to two reads (Fig. 4D). Next, we also examined miR-205 repression by the percent of miR-205 reads out of total CLEAR-CLIP reads per gene. We found that genes where miR-205 constituted >10% of the total CLEAR-CLIP reads were repressed whereas genes where miR-205 constituted 0%–10% of the total reads were not significantly repressed (Fig. 4E). Again, when miR-205 targets were binned into three groups based on reads per gene (one to two reads, three to nine reads, and 10+ reads) and then further divided into high and low percent miR-205 targeting (more or <10%), we found for each group the set of genes with a high percent of miR-205 reads was repressed better than the low percentage set (Fig. 4F). Finally, we also observed a strong correlation between the number of discrete CLEAR-CLIP sites per gene, especially for genes with greater than three sites and the number of sites with a seed match per gene on repression upon the induction of miR-205 (Supplemental Fig. S4A,B) or miR-205 (Supplemental Fig. S4C). Together, these analyses show that optimized CLEAR-CLIP quantitatively reflects the strength of miRNA-mediated regulation.

CLEAR-CLIP distinguishes miRNA 5′ isomiRs and their mRNA targets

Recently, miRNA isoforms that vary at the 5′ end or the 3′ end sequences were identified (Ameres and Zamore 2013; Guo and Chen 2014). Notably, variations at the 5′ end of miRNAs have the potential to change the miRNA seed and therefore mRNA target sites. To unequivocally identify miRNAs with 5′ variations and their associated targets requires single nucleotide resolution for both miRNAs and their targets, which is provided by CLEAR-CLIP data. First, we sought to determine whether the CLEAR-CLIP...
RNase step alters the detection of 5′ end of miRNAs. Due to the large number of possible isomiRs derived from murine miRNAs we first defined the isomiRs that exist in our system. To this end, we analyzed previously published small RNA sequencing data from skin epithelial cells without RNase treatment (Wang et al. 2012; Riemondy et al. 2015) and identified 2188 isomiRs derived from 447 miRNAs. The CLEAR-CLIP data were then remapped using the database of isomiRs. After mapping, we identified CLEAR-CLIP reads for 669 5′ isomiRs derived from 297 miRNAs that were abundantly detected. We observed the majority of reads for miR-200s and miR-205 starting at the canonical 5′ position; however, we also observed a significant number of reads starting one nucleotide downstream from the canonical 5′ end for miR-203 (Fig. 5A). The observed miR-203 isomiR patterns were previously confirmed (Zhang et al. 2013), validating the ability of CLEAR-CLIP to identify 5′ isomiRs.

We next calculated IsomiR:mRNA chimeric reads for each isomiR individually to identify isomiR-specific sites. For the majority of miRNAs, this analysis detected a clear preference of one 5′ isomiR over others (typically the canonical 5′ start position). For example, miR-200b with the canonical 5′ end start had 658 high-confidence target sites, whereas miR-200b with the 5′ end shifted downstream one nucleotide had only 11 high-confidence sites. Furthermore, nine of the 11 high-confidence areas observed in the miR-200b isomiR chimeras were also targeted by the canonical isomiR. Interestingly, when motif enrichment was performed on the high-confidence sites for miR-200b canonical isomiR versus the shifted isomiR, there was a corresponding shift in the motif (Fig. 5B).

miR-203 was the only miRNA in our system that had a significant number of high-confidence target sites for multiple 5′ isomiRs. We observed 248 sites for the canonical isomiR and 284 sites associated with the shifted isomiR (of these areas 58 of them overlapped). De novo motif searching in each group correctly identified the shift in the seed match motif, which would be expected for each isomiR (Fig. 5C, AUUUC vs. CAUUUC for the canonical and shifted isomiRs, respectively). Next, mRNA target sites associated only with canonical miR-203, only with the shifted miR-203 or shared by both were searched for the canonical 6mer, the shifted 6mer or the 7mer-m8 motif (which contains both the canonical and the shifted 6mer) (Fig. 5D). We observed a clear preference of each isomiR for their cognate motif (Fig. 5D–G) and areas shared by the isomiRs were more likely to contain a 7mer-m8 motif, as expected since it contains both the 6mer motifs. These data provide experimental evidence that two 5′ end isomiRs of miR-203 indeed utilize shifted seed sequences and target different mRNAs. Taken together, these results highlight the ability of CLEAR-CLIP to distinguish 5′ end isomiRs and their targets at single nucleotide resolution.

Comparison between CLEAR-CLIP identified targets and TargetScan predicted targets

CLEAR-CLIP and computational algorithms such as TargetScan are two different approaches that can provide miRNA- and site-specific information for miRNA targeting. We therefore compared the performance of our CLEAR-CLIP method and TargetScan predictions for mouse (TargetScan 7.1 mouse using evolutionarily conserved miRNA sites). Because CLEAR-CLIP identifies targets which are expressed in a cellular context and are bound by miRNAs regardless of whether they harbor seed matches, and TargetScan predicts only targets with a seed independently of gene expression, we only used TargetScan predicted genes that were expressed in our system (base mean >10 in RNA-seq) and required our CLEAR-CLIP targets to have a 7mer or 8mer seed match in the 3′ UTR, identical to target selection criteria we used from TargetScan. Pooling together both miR-200 seed types, we identified 436 expressed genes that were shared between our CLEAR-CLIP data and TargetScan predictions (Fig. 6A). We also found 366 genes that were only identified in our CLEAR-CLIP and 854 genes that were only predicted by TargetScan. To determine the effectiveness of these miR-200 targets, we examined the repression of mRNA abundance of CLEAR-CLIP targets versus TargetScan predictions upon miR-200 induction. Overall, CLEAR-CLIP targets were better repressed upon miR-200 induction than genes predicted by TargetScan (Fig. 6B). Examining the overlap between these groups we found that genes in our CLEAR-CLIP and predicted by TargetScan were best repressed by miR-200 induction, followed by genes only identified by CLEAR-CLIP and then genes only predicted by TargetScan (Fig. 6C).

To determine whether gene expression contributes to the lack of detection by CLEAR-CLIP, we examined the expression level in our RNA-seq data of genes found in the different overlaps. Genes predicted only by TargetScan and not found in our CLEAR-CLIP did have slightly lower expression levels, but many of the genes only predicted by TargetScan were expressed at a similar level to genes detected by CLEAR-CLIP (Supplemental Fig. S4D), suggesting that the lack of detection was not simply due to low expression levels. Furthermore, although many common targets between CLEAR-CLIP and TargetScan predictions were indeed based on the exact same sites, sometimes a gene was found by CLEAR-CLIP and TargetScan, but targeted at different sites within the 3′ UTR. For example, Brd4 contained one site that was captured by CLEAR-CLIP and also predicted by TargetScan, but it also had one site with a seed only captured by CLEAR-CLIP, and a TargetScan predicted site that did not have any CLEAR-CLIP reads (Fig. 6D). Ammecr1l contained one robust CLEAR-CLIP site with a seed that was not predicted by TargetScan and a predicted site that did not
have any miR-200 CLEAR-CLIP reads (Fig. 6E). Finally, TargetScan misses any sites that lack a canonical seed. For example, we found a heavily targeted site in the 3′ UTR of Tnrc6a that lacked a canonical seed and instead had a seed match with a G:U wobble (GGTATT instead of AGTATT) (Fig. 6F). Consistent with the binding data, Tnrc6a was repressed by 20% upon induction of miR-200.

To extend the study beyond miR-200s, we performed a similar comparison for targets of miR-205 captured by CLEAR-CLIP and predicted by TargetScan. Possibly due to a lower percentage of miR-205 sites containing a 7mer or 8mer seed, there was even less overlap between CLEAR-CLIP and TargetScan predictions, with only 59 genes identified by both. TargetScan predicted 320 genes
that were not identified by CLEAR-CLIP and our CLEAR-CLIP identified 104 genes that were not predicted by TargetScan (Fig. 6G). Despite identifying fewer targets, we still found that CLEAR-CLIP targets were repressed at the mRNA level more effectively than TargetScan predictions upon induction of miR-205 (Fig. 6H). Furthermore, the 104 genes only detected by CLEAR-CLIP were repressed similarly to the 59 common targets. The 320 genes only predicted by TargetScan were minimally repressed (Fig. 6I).

Next, we performed a global comparison of CLEAR-CLIP and TargetScan predictions, again using only targets we identified in CLEAR-CLIP with a 7mer or 8mer in the 3′ UTR and limiting TargetScan predictions to expressed genes (Fig. 6J).
genes. All miRNAs with at least 50 high-confidence target sites were used to analyze their predicted versus captured sites. We found 4992 miRNA:mRNA interactions that were shared between our CLEAR-CLIP and TargetScan predictions. However, 34,452 interactions were predicted by TargetScan but not captured by CLEAR-CLIP and 4230 interactions were captured by CLEAR-CLIP but not predicted by TargetScan (Fig. 6J). These results indicate differences in the methods and highlight the value of direct identification of targets for narrowing down miRNA recognized genes over prediction algorithms.

CLEAR-CLIP allows genome-wide discovery of miRNA-regulated gene networks

We next examined global features of miRNA-mediated target recognition by examining abundantly expressed miRNAs and their targets. To accomplish this, we used all miRNAs that had 50 or more high-confidence target sites, of which there were 88 miRNAs. First, we calculated high-confidence sites for each miRNA that were found in 2+, 5+, or 8+ libraries and then annotated these areas to the genome. Similar to the pattern of miR-200s (Fig. 1F), miRNA target sites were highly enriched in 3’ UTRs (Fig. 7A), especially when requiring sites to be found in more libraries. To map the global binding preference along the length of these 88 miRNAs, we calculated the fractional binding for each nucleotide using RNAhybrid across all high-confidence sites. The majority of miRNAs appear to rely heavily on seed binding, but the dependence on seed versus 3’ end sequences varies by miRNAs (Fig. 7B). Across these miRNAs, the seed nucleotides (2–8) are bound in more than 75% of occurrences. The fraction bound decreases at nucleotides 9 and 10, then increases for nucleotides 11 through 15 and trails off toward the 3’ end of the miRNA (Fig. 7C).

We next examined the presence of seed matches within captured mRNA target sites of the top 40 miRNAs. For these miRNAs, we searched for enriched 8 nucleotide sequences in high-confidence sites in 3’ UTRs using HOMER (Heinz et al. 2010). For most miRNAs, perfect matches to 6mer seed sequences were the most enriched motif (Fig. 7D). However, a few miRNAs also have some slight variation of the 6mer sequences mostly corresponding to nucleotides 2 or 7, particularly miR-21-5p, miR-125a-5p, and miR-203-3p (possibly due to the mixture of miR-203 isomiRs as shown above). In addition, the most prolific miRNA, miR-31-5p, binds to 2145 sites whereas the least prolific miRNA among the top 40 miRNAs, let-7d-5p, binds to 202 sites (Fig. 7D). These data demonstrate that a miRNA can robustly interact with hundreds to thousands of target sites in 3’ UTRs even in one cell type. Interestingly, for each of these top 40 miRNAs, we observed 35.5%–67.2% of miRNA-associated miRNAs contain the seed match found by HOMER. Because seedless miRNAs have a minimal impact on gene expression for both miR-200s and miR-205 (Fig. 2D,E,G), these data may indicate a general activity of target scanning by many miRNAs that are captured by CLEAR-CLIP.

miRNA-mediated gene expression regulation is highly complex. To date, it remains unclear how many genes and pathways are regulated by miRNAs in a specific cellular context. To gain insights into the overall function of the miRNA pathway in epithelial cells of the skin, we analyzed the targets of these abundantly expressed miRNAs to determine if different miRNAs work in concert to regulate similar genes or pathways and which genes and signaling pathways are most heavily affected by the miRNA pathway. To accomplish this, we used the same miRNAs with 50 or more high-confidence sites as above, calculated high-confidence sites within 3’ UTRs with a seed match (7mer or better) and determined gene lists for each miRNA. We then performed hierarchical clustering on this set of miRNAs and their targets. As shown in Supplemental Figure S5, miRNAs that clustered together by genes targeted mostly shared identical seeds, indicating that the identical seed match is the strongest driver for miRNA target coordination. Next, we used Enrichr (Kuleshov et al. 2016) to classify gene lists into GO terms by KEGG pathways and analyzed whether different miRNAs regulate similar cellular functions. We found that miRNA families tend to cluster together such as a cluster of let-7 miRNAs (Fig. 7E). However, we observed miR-30b-5p also clustered together with the let-7 family. Clustering of miR-15a/b, miR-16, and miR-29a/b was also seen, which is noteworthy as these miRNAs have only one base pair difference in their seed regions and indeed appear to be regulating the same pathways in a similar fashion to the miR-200 family. Interestingly, the miR-200 family, miR-19a, miR-203, miR-301, and miR-27a/b also form a cluster, raising the possibility that different miRNAs coordinately target similar pathways.

To better determine what pathways are strongly targeted by the miRNA pathway globally, we calculated how many miRNAs were targeting each GO term. This analysis revealed that the majority of the top categories targeted by miRNAs in epithelial cells of the skin are important regulators of cancer (Fig. 7F). In particular, PI3K-Akt signaling, focal adhesion, Hippo pathway, and p53 pathways were strongly targeted.

Because the number of CLEAR-CLIP reads reflects the strength of repression (Fig. 4), we next identified strongly regulated targets by all miRNAs. We calculated the total number of unique miRNAs targeting each gene and the total number of CLEAR-CLIP reads per gene. To identify highly targeted genes, we selected for genes that were targeted by 4+ different miRNAs and harbored 40 or more total CLEAR-CLIP reads. To test the effectiveness of this approach, we utilized RNA-seq data from wild-type versus Exportin-5 (Xpo5) knock out mouse epidermis, in which
most miRNAs are depleted. We then intersected the Xpo5 RNA-seq with highly targeted genes determined by CLEAR-CLIP. Genes with 4+ miRNAs and 40 or more total miRNA reads were more strongly up-regulated at the mRNA level than all genes targeted by miRNAs, and both groups of genes were derepressed relative to nontargeted genes upon the depletion of miRNAs (Fig. 7G).

These data validate that highly targeted miRNAs determined by CLEAR-CLIP are indeed under strong regulation by the miRNA pathway.

Finally, we performed GO analysis for KEGG terms on genes heavily targeted by the miRNA pathway.
(Supplemental Table S2). Many of the top pathways were again related to tumorigenesis. The top category was Proteoglycans in Cancer, with 52 out of 203 genes in this category heavily targeted by the miRNA pathway. Three of the top seven categories were also related to cell adhesion including focal adhesion, adherens junction, and regulation of actin cytoskeleton. Taken together, these data offer new insights about miRNA-mediated regulation in epithelial cells of the skin and provide a molecular basis to explore the miRNA pathway as an important negative regulator of tumorigenesis. Because CLEAR-CLIP should be generally applicable to all cell types, we envision this method can be broadly used to identify all miRNA and mRNA interactions in a cellular context-specific manner.

DISCUSSION

In this study, we have directly captured miRNA:RNA interactions at the genomic scale and, as a proof-of-principle, probed the action of miRNAs in wild-type, miR-200 family knockout and induced epithelial cells. The use of PEG-8000, randomized 3’ and 5’ adapters (Zhang et al. 2013) and enhanced RNA isolation from membrane (Fig. 1A; Zarnegar et al. 2016) improves the quantitative performance of CLEAR-CLIP. As a result of these improvements, the number of CLEAR-CLIP reads (Fig. 4A,D), the percentage of individual miRNA CLEAR-CLIP reads among the total CLEAR-CLIP reads on the same gene (Fig. 4B,C,E,F) and the number of unique CLEAR-CLIP identified miRNA binding sites (Supplemental Fig. S4A–C) can be used to measure the strength of miRNA-mediated regulation. These analyses are also supported by our recent study that individual sites harboring more CLEAR-CLIP reads generally confer stronger regulation as assayed by the classic luciferase assay (Hoefert et al. 2018). CLEAR-CLIP also allowed measurement of binding by other miRNAs upon the perturbation of miR-200s. An increase in miR-205 binding was seen upon DKO of miR-200s; however, no change was seen upon induction of miR-200s (Fig. 1E). It is unclear why only the DKO of miR-200s caused a shift in miR-205 binding. It is possibly due to differences between the cell lines and their Ago expression levels; however, Ago protein levels are tightly regulated by miRNA levels (Wang et al. 2012; Smibert et al. 2013). The quantitative performance of our CLEAR-CLIP is further validated by the detection of more CLEAR-CLIP reads on the same sites in miR-200 induced epithelial cells than control cells (Fig. 2A,C). Importantly, genes commonly targeted in control and miR-200 induced cells are more strongly down-regulated than genes uniquely bound by miR-200s in the induced cells (Fig. 2B). These data suggest that elevated miRNA expression preferentially regulates existing targets rather than de novo targets, and this observation should facilitate the study of other miRNAs with gain-of-function approaches.

Direct and quantitative capture of miRNA targets also provides new insights into how miRNAs recognize their targets. Although perfect seed matches, in particular 7mer and 8mer matches, result in the strongest regulation (Fig. 2D–H), miRNAs also reproducibly bind to a large number of sites that lack a seed match (Fig. 7D; Supplemental Fig. S2E). These observations suggest that miRNAs and their associated RISC may scan a large number of sites within and outside of 3’ UTRs, perhaps through an AGO2 phosphorylation-dependent mechanism that has been recently demonstrated (Golden et al. 2017). The widespread, scanning activities of miRNAs suggest a possibility that a large portion of the transcriptome is under surveillance of miRNAs but only the best targets are actively down-regulated.

Although miRNAs predominantly use their 5’ seed regions to recognize their functional targets, base-pairing at the 3’ regions of individual miRNAs has also been previously described (Chi et al. 2009; Schirle et al. 2014; Moore et al. 2015; Bartel 2018). Analysis of top miRNAs in epithelial cells reveals that different miRNAs have different preferences to matches in their 3’ regions (Fig. 7B,C). For example, miR-200 miRNAs have two short 3’ regions that help target recognition. In particular, nucleotides 12–14, which are identical in all family members (GGU), show a strong preference to base pair with their targets, providing an explanation for the overlapping targets of different miR-200s with identical seed region (Fig. 3C). In contrast, miR-205 prefers 3’ end binding of nucleotides 15–20 (Fig. 3D).

This study also examined isomiRs and identified their associated mRNA sites in an isomiR-specific manner (Fig. 5). Our data shows that 5’ end isomiRs can be reliably identified in CLEAR-CLIP, due to the protection of the 5’ end of miRNAs by the MID domain of Ago protein (Schirle et al. 2014). While all 5’ end isomiRs that we could identify were examined, we found that only miR-203 produced significant amounts of 5’ isomiRs. Importantly, we were able to identify differences in mRNA targeting by the two isomiRs of miR-203. This analysis suggests that CLEAR-CLIP distinguishes 5’ end isomiRs and their targets at single-nucleotide resolution. The above comparison between CLEAR-CLIP identified and TargetScan predicted targets (Fig. 6) offers further validation of using an experimental approach. In particular, the ability to reduce many false positive predictions should help detection of highly regulated targets and their relevant pathways.
cancer, focal adhesion, adherens junction, FoxO signaling, and Hippo signaling among others (Fig. 7F; Supplemental Table S2). These results provide support to published reports indicating that miRNA dysregulation is causal in many types of cancer (Lu et al. 2005; Jansson and Lund 2012; Rupaimoole and Slack 2017). These data are also consistent with genetic studies of Dicer1, Dgcr8, and Ago1/2, in which deletion of these essential factors of the miRNA pathway in epithelial cells of the skin does not change cell fate but leads to defects in hair morphogenesis and stem cell maintenance (Andl et al. 2006; Yi et al. 2006; Wang et al. 2012). Although the detailed mechanisms remain to be elucidated, the comprehensive mapping of the miRNA targeted transcriptome will lay a foundation to answer the question of how miRNAs regulate hair morphogenesis and skin development.

MATERIALS AND METHODS

CLEAR-CLIP

Detailed CLEAR-CLIP protocol is included in the Supplemental Information. Bioinformatic analysis for assigning chimeric reads and genome annotation for CLEAR-CLIP was done as described previously (Hoefert et al. 2018).

Unbiased motif finding using Homer

Homer was downloaded from http://homer.ucsd.edu/homer/motif/. High-confidence areas found in 2+ libraries were calculated for the top 40 miRNAs (using BedTools MultiIntersectBed), and selected for areas overlapping 3’UTRs using Bedtools Intersect. All 3’UTR areas for each miRNA were then processed through Homer. The background for motif finding was all mouse mm10 3’UTRs. Settings for motif finding were: -len 8 -size given -rna -noweight -p 2 -chopify. The top ranked motif is shown for each miRNA.

GO terms clustering

GO terms were calculated using Enrichr (https://amp.pharm.mssm.edu/Enrichr/). The KEGG 2016 categories were used for GO Terms and the combined score was used as output for graphing. Scores were then hierarchical clustered using Gene Cluster 3.0 (https://www.encodeproject.org/software/cluster/) and drawn as a heat map using Java Tree View (http://jtreeview.sourceforge.net/).

qPCR

miRNA qPCR was performed using the Qiagen miScript II RT Kit (#218160). miRNAs were quantified from the cDNA using iQ SYBR green supermix (170-8880; Bio-Rad Laboratories) and the ΔΔC(t) method relative to U6 RNA. Forward primers for miRNAs are as follows: miR-205 5′ TCTTCTATTCCACCGGAGTCTG 3′, miR-200a 5′ TAACACTGTCGGTAAATGATGA 3′, miR-429 5′ TAATAGTGTTCCAAACGTGG 3′, miR-141 5′ TAATACTGTCGGTAAATGATGGA 3′, miR-141 5′ TAATACTGTCGGTAAATGATGGA 3′. Reverse primer used was Qiagen’s miScript Universal Primer.

RNA-seq

RNA-seq was performed on both miR-200b cluster induced cells and miR-205 induced cells, each relative to the cells only treated with vehicle. Library preparation was performed using the NEBNext Ultra Directional RNA Library Prep Kit for Illumina (#E7420). Sequencing reads were mapped to the mm10 genome using Bowtie2 and gene counting was performed using HTSeq-count. Differential analysis was performed using DESeq2. Xpo5 RNA-seq was performed as above using Xpo5 knock out whole epidermis versus control epidermis.

TargetScan comparisons

All comparisons were done to TargetScan Mouse 7.1, using their database of conserved sites. These were compared to CLEAR-CLIP data for genes with a high-confidence site (the same miRNA area found in two or more libraries ligated to the same miRNA) that contained an 8mer or either 7mer (M8 or A1). Comparisons were done by gene, irrespective of whether the identical site was found in both. Comparisons were also only done for genes that had a base mean >10 in our RNA-seq data to make sure we compared genes that were expressed in our system.

IsomiR detection and CLEAR-CLIP mapping of isomiRs

We defined isomiRs that are expressed in skin epithelial cells using small RNA sequencing from studies previously published by our laboratory (Wang et al. 2012; Riendy et al. 2015). This was accomplished by BLAST mapping small RNA sequencing reads to a database of miRNA hairpins downloaded from miRbase.org. The start and end position of where the read mapped on the hairpin to was then used to define an individual isomiR. We analyzed all miRNA with 100 or more CLEAR-CLIP reads from our 18 libraries. Any isomiR that was found to constitute >1% of the total reads for that miRNA was then used in a database to map CLEAR-CLIP reads to. CLEAR-CLIP reads were remapped as previously described (Hoefert et al. 2018) using the database of isomiRs found and reads 3’ downstream from isomiRs (presumed mRNAs) were remapped again using Novoalign. Further analyses were performed only taking into account the 5’ end position of the isomiR. Defining high-confidence areas and HOMER motif finding parameters are described above in the methods.

Statistics

All measurements were taken from separate samples. All statistical tests used are described in the figure legends. For graphing of log2 data, such as Figure 2A, one was added to each data point to make the log2 of zero possible.
DATA DEPOSITION

Most analyses were performed using publicly available programs such as BEDTools. Custom scripts such as CLEAR-CLIP mapping steps, Bedfile area extending, BLAST output processing, and scripts to parse RNAhybrid are available at https://github.com/Bjerkega/CLEAR-CLIP-analysis-scripts. High-throughput sequencing data are available online from the Gene Expression Omnibus (GEO) at GSE102716 (CLEAR-CLIP, including raw sequencing files and processed reads of miRNA:mRNA chimeras) and GSE131205 (RNA-seq). Mapped CLEAR-CLIP reads can also be visualized at https://genome.ucsc.edu/s/Bjerkega/Bjerkega%20CLEAR%20DCLIP.

SUPPLEMENTAL MATERIAL

Supplemental material is available for this article.

ACKNOWLEDGMENTS

We thank K. Diener and B. Gao for sequencing; J. Hoeffert and D. Wang for generating miR-200 and miR-205 mouse models and isolating primary cells; J. Lee for generating XPOS RNA-seq data sets; K. Riemann for helping with bioinformatics; and members of the Yi laboratory for discussion. Research reported in this publication was supported by the National Institute of Arthritis and Musculoskeletal and Skin Diseases of the National Institutes of Health under award numbers R01AR059697 and R01AR066703 (to R.Y.). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. G.A.B. was supported by an American Cancer Society postdoctoral fellowship (129540-PF-16-059-01-RMC).

Received October 9, 2019; accepted December 27, 2019.

REFERENCES

Ambros V. 2004. The functions of animal microRNAs. Nature 431:350–355. doi:10.1038/nature02871

Ameres SL, Zamore PD. 2013. Diversifying microRNA sequence and function. Nat Rev Mol Cell Biol 14:475–488. doi:10.1038/nrm3611

Andl T, Murchison EP, Liu F, Zhang Y, Yunta-Gonzalez M, Tobias JW, Andl CD, Seykora JT, Hannon GJ, Millar SE. 2006. The miRNA-processing enzyme dicer is essential for the morphogenesis and maintenance of hair follicles. Curr Biol 16:1041–1049. doi:10.1016/j.cub.2006.04.005

Baek D, Villen J, Shin C, Camargo FD, Gygi SP, Bartel DP. 2008. The impact of microRNAs on protein output. Nature 455:64–71. doi:10.1038/nature07242

Bartel DP. 2009. MicroRNAs: target recognition and regulatory functions. Cell 136:215–233. doi:10.1016/j.cell.2009.01.002

Bartel DP. 2018. Metazoan microRNAs. Cell 173:20–51. doi:10.1016/j.cell.2018.03.006

Baumjohann D, Ansel KM. 2013. MicroRNA-mediated regulation of T helper cell differentiation and plasticity. Nat Rev Immunol 13:666–678. doi:10.1038/nri3494

Bleazard T, Lamb JA, Griffiths-Jones S. 2015. Bias in microRNA functional enrichment analysis. Bioinformatics 31:1592–1598. doi:10.1093/bioinformatics/btv203

Broughton JP, Lovci MT, Huang JL, Yeo GW, Pasquinelli AE. 2016. Pairing beyond the seed supports microRNA targeting specificity. Mol Cell 64:320–333. doi:10.1016/j.molcel.2016.09.004

Chi SW, Zang JB, Mele A, Damell RB. 2009. Ago HITS-CLIP decodes miRNA-mRNA interaction maps. Nature 460:479–486. doi:10.1038/nature08170

Fang S, Zhang L, Guo J, Niu Y, Wu Y, Li H, Zhao L, Li X, Teng X, Sun X, et al. 2018. NONCODEv5: a comprehensive annotation database for long non-coding RNAs. Nucleic Acids Res 46:D308–D314. doi:10.1093/nar/gkx1107

Golden RJ, Chen B, Li T, Braun J, Manjunath H, Chen X, Wu J, Schmid V, Chang T, Kopf F, et al. 2017. An Argonaute phosphorylation cycle promotes microRNA-mediated silencing. Nature 542:197–202. doi:10.1038/nature21025

Guo L, Chen F. 2014. A challenge for miRNA: multiple isomiRs in miRNomics. Gene 544:1–7. doi:10.1016/j.gene.2014.04.039

Guo H, Ingolia NT, Weissman JS, Bartel DP. 2010. Mammalian microRNAs predominantly act to decrease target mRNA levels. Nature 466:835–840. doi:10.1038/nature09267

Gurtan AM, Sharp PA. 2013. The role of miRNAs in regulating gene expression networks. J Mol Biol 425:3582–3600. doi:10.1016/j.jmb.2013.03.007

Hafner M, Landthalder M, Burger L, Khorshid M, Hauser J, Beminger P, Rothbiller A, Ascano M, Jungkamp A-C, Munschauer M, et al. 2010. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. Cell 141:129–141. doi:10.1016/j.cell.2010.03.009

Heinz S, Bannet C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. 2010. Simple combinations of lineages-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell 38:576–589. doi:10.1016/j.molcel.2010.05.004

Helwak A, Kudla G, Dudnakova T, Tollervey D. 2013. Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding. Cell 153:654–665. doi:10.1016/j.cell.2013.03.043

Hoeffert JE, Bjerke GA, Wang D, Yi R. 2018. The microRNA-200 family coordinately regulates cell adhesion and proliferation in hair morphogenesis. J Cell Biol 217:2185–2204. doi:10.1083/jcb .201708173

Jansson MD, Lund AH. 2012. MicroRNA and cancer. Mol Oncol 6:590–610. doi:10.1016/j.molonc.2012.09.006

Jons S, Izaurralde E. 2015. Towards a molecular understanding of microRNA-mediated gene silencing. Nat Rev Genet 16:421–433. doi:10.1038/nrg3963

Kudla G, Granneman S, Hahn D, Beggs JD, Tollervey D. 2011. Cross-linking, ligation, and sequencing of hybrids reveals RNA-RNA interactions in yeast. Proc Natl Acad Sci USA 108:10010–10015. doi:10.1073/pnas.101786108

Kuleshov MV, Jones MR, Roulillard AD, Fernandez NF, Duan Q, Wang Z, Kolev S, Jenkins SL, Jagodnik KM, Lachmann A, et al. 2016. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res 44:W90–W97. doi:10.1093/nar/gkw377

Lin S, Gregory RI. 2015. MicroRNA biogenesis pathways in cancer. Nat Rev Cancer 15:321–333. doi:10.1038/nrc3932

Lu J, Getz G, Miska EA, Alvarez-Saavedra E, Lamb J, Peck D, Sweet-Cordero A, Ebert BL, Mak RH, Ferrando AA, et al. 2005. MicroRNA expression profiles classify human cancers. Nature 435:834–838. doi:10.1038/nature03702

Mendell JT, Olson EN. 2012. MicroRNAs in stress signaling and human disease. Cell 148:1172–1187. doi:10.1016/j.cell.2012.02.005

Moore MJ, Scheel TKH, Luna J, Park CY, Fak JJ, Nishiuchi E, Rice CM, Damell RB. 2015. miRNA-target chimera reveal miRNA ’ end pairing as a major determinant of Argonaute target specificity. Nat Commun 6:1–17. doi:10.1038/ncomms9864
Pinzón N, Li B, Martinez L, Sergeeva A, Presumey J, Apparailly F, Seitz H. 2017. MicroRNA target prediction programs predict many false positives. Genome Res 27: 234–245. doi:10.1101/gr.205146.116

Rehmsmeier M, Steffen P, Hochsmann M, Giegerich R. 2004. Fast and effective prediction of microRNA/target duplexes. RNA 10: 1507–1517. doi:10.1261/rna.5248604

Riemondy K, Wang XJ, Torchia EC, Roop DR, Yi R. 2015. MicroRNA-203 represses selection and expansion of oncogenic Hras transformed tumor initiating cells. Elife 4: 1–26. doi:10.7554/elife.07004

Rupaimoole R, Slack FJ. 2017. MicroRNA therapeutics: towards a new era for the management of cancer and other diseases. Nat Rev Drug Discov 16: 203–222. doi:10.1038/nrd.2016.246

Saito T, Sætrom P. 2012. Target gene expression levels and competition between transfected and endogenous microRNAs are strong confounding factors in microRNA high-throughput experiments. Silence 3: 3. doi:10.1186/1758-907X-3-3

Schirle NT, Sheu-Gruttadauria J, Macrae IJ. 2014. Structural basis for microRNA targeting. Science 346: 608–613. doi:10.1126/science.1258040

Smibert P, Yang J-S, Azzam G, Liu J-L, Lai EC. 2013. Homeostatic control of Argonaute stability by microRNA availability. Nat Struct Mol Biol 20: 789–795. doi:10.1038/nsmb.2606

Thomson DW, Bracken CP, Goodall GJ. 2011. Experimental strategies for microRNA target identification. Nucleic Acids Res 39: 6845–6853. doi:10.1093/nar/gkr330

Wang D, Zhang Z, O’Loughlin E, Lee T, Houel S, O’Carroll D, Tarakhovsky A, Ahn NG, Yi R. 2012. Quantitative functions of Argonaute proteins in mammalian development. Genes Dev 26: 693–704. doi:10.1101/gad.182758.111

Wang D, Zhang Z, O’Loughlin E, Wang L, Fan X, Lai EC, Yi R. 2013. MicroRNA-205 controls neonatal expansion of skin stem cells by modulating the PI3K pathway. Nat Cell Biol 15: 1153–1163. doi:10.1038/ncb2827

Yi R, O’Carroll D, Pasolli HA, Zhang Z, Dietrich FS, Tarakhovsky A, Fuchs E. 2006. Morphogenesis in skin is governed by discrete sets of differentially expressed microRNAs. Nat Genet 38: 356–362. doi:10.1038/ng1744

Zarnegar BJ, Flynn RA, Shen Y, Do BT, Chang HY, Khavari PA. 2016. irCLIP platform for efficient characterization of protein-RNA interactions. Nat Methods 13: 489–492. doi:10.1038/nmeth.3840

Zhang Z, Lee JE, Riemondy K, Anderson EM, Yi R. 2013. High-efficiency RNA cloning enables accurate quantification of miRNA expression by deep sequencing. Genome Biol 14: R109. doi:10.1186/gb-2013-14-10-r109