scientific reports

OPEN

New characteristics of MiRNA and IsomiR interactions with mRNA

Matthew Weston[1], Rony Chowdhury Ripan[1], Xiaoman Li[2✉] & Haiyan Hu[1✉]

Studying the interactions between microRNAs/isomiRs and mRNAs is crucial due to their fundamental roles in gene regulation and disease. Although many isomiRs have been identified, the analysis of their interactions with mRNAs remains in its early stages. In this study, we compiled available human chimeric reads, each pairing a microRNA or isomiR segment with an mRNA segment. We then identified 1747 isomiRs and over 5 million microRNA/isomiR–mRNA interactions from the complied data. We found that microRNAs with higher adenine and thymine content, and lower cytosine content, tend to have more isomiRs and target more mRNAs. Notably, 18.9% of mRNA targets were bound exclusively by isomiRs, not their microRNAs. Furthermore, isomiRs sharing the same seed sequences as their reference miRNAs may target different mRNAs from their miRNAs, suggesting functional divergence. Interestingly, 20.0% of microRNAs and 8.2% of isomiRs bind mRNAs independently of their seed regions. Among those that do utilize seed regions, 94.5% of microRNAs and 95.7% of isomiRs also engage non-seed region binding. Our findings provide new insights into the complexity of microRNA/isomiR–mRNA interactions.

IsomiRs are isoforms of reference microRNAs (miRNAs). They arise from alternative miRNA biogenesis processes such as nucleotide additions, deletions, or substitutions. These changes often result in differences in length and sequence compared with their reference miRNAs[1]. Like miRNAs, isomiRs regulate gene expression by binding to mRNA targets, leading to mRNA degradation or translational repression[2]. However, due to variations in their seed regions, isomiRs may target different mRNAs from their reference miRNAs, affecting gene expression in both normal and disease conditions[3]. It is thus important to study isomiRs, especially their targeting mechanisms.

Previous studies have classified isomiRs into four major types: 5′ isomiRs, 3′ isomiRs, polymorphic isomiRs, and mixed type isomiRs[4,5]. The first two types differ in length from their reference miRNAs, while the third type has the same length. The mixed type has both length and nucleotide variations. Research has also explored how isomiRs with altered seed regions target different mRNAs compared with their reference miRNAs[2,3,6]. They have also investigated the potential of isomiRs as biomarkers for disease detection, prognosis, and diagnosis[7–10]. In addition, several databases have been developed to store and organize isomiR-related data, such as isomiRdb, IsomiR Bank and the tumor isomiR encyclopedia[11–13]. For instance, isomiRdb has cataloged over 70,000 isomiRs across approximately 43,000 samples[11].

Despite these efforts, only a few studies have systematically examined isomiR–mRNA target interactions. One such study, isomiRTar[14], analyzed the predicted targets of 1022 5′ isomiRs across 31 cancer types using tools such as miRDB[15] and targetScan[16]. However, these tools were originally designed for miRNAs and may not be optimal for isomiR targeting. Another effort, DMISO[17], trained a deep learning model on crosslinking, ligation and sequencing of hybrids (CLASH) data to predict miRNA/isomiR–mRNA interactions[18]. CLASH captures direct miRNA/isomiR–mRNA interactions by generating chimeric reads composed of physically ligated miRNA/isomiR fragments and their mRNA target segments, providing context-specific interaction evidence. While DMISO is a step forward, it was trained on only six CLASH samples, limiting its generalizability. Improving prediction models requires both more data and a deeper understanding of miRNA/isomiR binding mechanisms[17,19–22].

In the past decade, additional CLASH-like technologies such as CLEAR-CLIP, Quick CLASH (qCLASH), and chimeric e-CLIP have emerged. These methods generate high-resolution chimeric reads of miRNA/isomiR–mRNA interactions[23–25]. These additional data offer an unprecedented opportunity to systematically investigate miRNA/isomiR targeting across conditions and cell types.

[1]Department of Computer Science, University of Central Florida, Orlando, FL, USA. [2]Burnett School of Biomedical Sciences, College of Medicine, University of Central Florida, Orlando, FL, USA. ✉email: xiaoman@mail.ucf.edu; haihu@cs.ucf.edu

In this study, we analyzed 1747 isomiRs and 5,556,801 miRNA/isomiR–mRNA interactions derived from 67 publicly available CLASH-like human samples across six studies. Of the isomiRs identified, 73.1% were already cataloged in isomiRdb[11], while the remaining 26.9% likely represent novel isomiRs. Approximately 6.7% of the inferred miRNA/isomiR–mRNA interactions were recorded in miRTarBase (Version 10.0)[26]. Notably, we found that miRNAs with more isomiRs and mRNA targets tend to have higher adenine and thymine content and lower cytosine content. These findings suggest novel sequence characteristics associated with broader target engagement in miRNAs and isomiRs.

## Materials and methods
Process CLASH and CLASH-like data
We obtained 67 human CLASH and CLASH-like samples across six datasets (Supplementary Tables S1, S2), the only available samples with chimeric reads we could access. These included nine samples from CLASH[18], twelve from CLEAR-CLIP[25], nine from qCLASH[23], thirteen from chimeric eCLIP[24], three from Kozer et al.[27], and twenty-one from Fields et al.[28].

For each sample, we followed the DMISO[17] protocol to process reads and extract miRNA/isomiR–mRNA interactions. The DMISO protocol closely mirrored the methodology used in the original CLASH study[18] (Fig. 1). We first removed adaptors in reads and filtered low-quality reads using trimmomatic[29] and fastqc. Next, we eliminated duplicate reads with the ShortRead[12] library in R. The processed reads were then mapped to GENCODE[30] protein-coding transcript sequences (version 46) and miRBase[31] human mature miRNA sequences (version 22.1) using BLAST[32] (version 2.15.0+). Mapping was performed with an E-value threshold of 0.1 and a default word size of 11. Reads mapped to the antisense strand or containing alignment gaps were excluded as in the original study[18].

Each read mapped to both mRNA and miRNA was considered a chimeric read. Because miRNA portions are typically shorter and more prone to sequencing errors, we retained only chimeric reads that had the same miRNA segments as at least nine other chimeric reads. We also excluded chimeric reads with more than a four-nucleotides gap or overlap between the mRNA and miRNA portions. Since both miRNA and mRNA segments could map to multiple miRNA and mRNA transcripts, we applied the following criteria to select the most significant miRNA–mRNA pair: (1) the pair with the lowest BLAST E-values, and (2) if e-values were identical, the pair with the highest BLAST bit score. In total, we identified 5,556,801 chimeric reads across the 67 samples from the six datasets (Supplementary Table S2, Fig. 1).

Identify IsomiRs and miRNA/isomiR–mRNA interactions
Following the procedure outlined above, each chimeric read contains two segments: a mRNA segment and a miRNA segment, with at most four nucleotides of overlap between them. We classify the miRNA segment as an miRNA if it is an exact substring of a mature miRNA sequence in miRbase[31] (Version 22.1). If it does not perfectly match a mature miRNA, we designate it as an isomiR, provided this exact miRNA segment is present in at least ten chimeric reads. This ten-read threshold ensures that the probability of a BLAST hit arising by chance is below 8.9E-8, consistent with the cutoff used in DMISO[17]. Unlike DMISO, when a miRNA segment
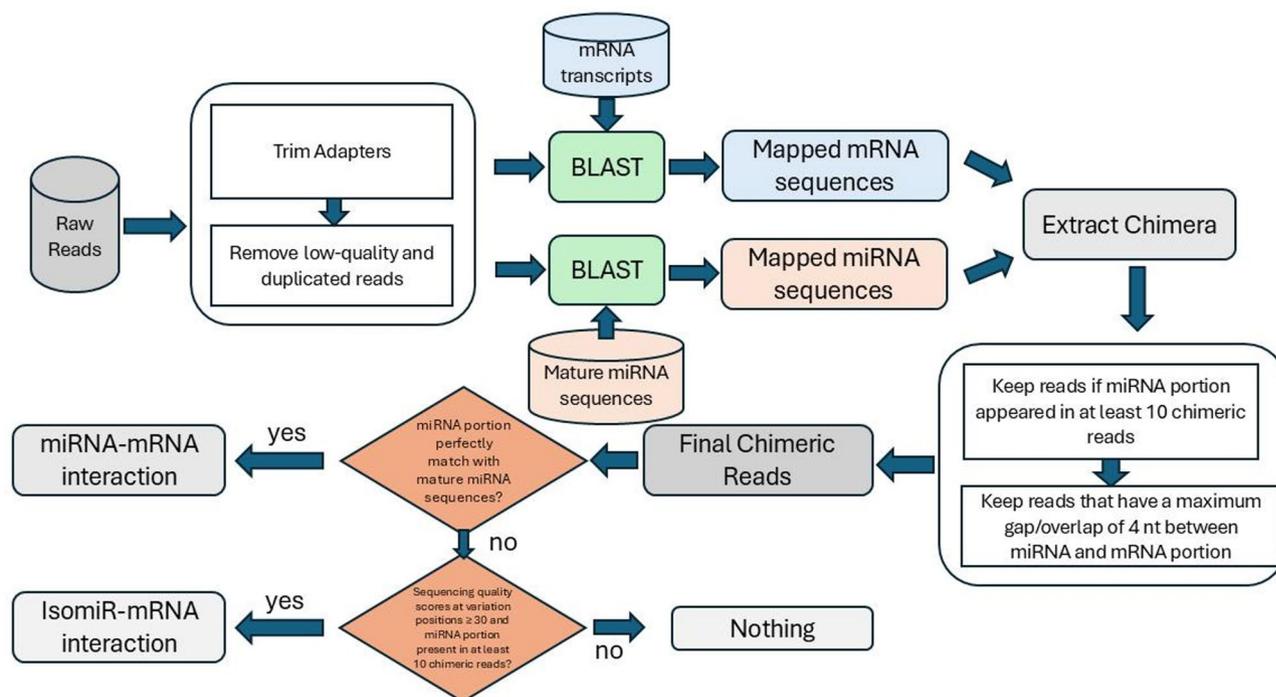
Fig. 1. The pipeline for extracting miRNA/isomiR–mRNA interactions.

only partially matches a mature miRNA, we extend it to the full-length by appending the missing portion of the mature miRNA. Because CLASH and CLASH-like experiments cannot tell whether a miRNA segment is from a full-length miRNA/isomiR or a truncated form generated during biogenesis (e.g., a 3′ isomiR), we do not use the extended miRNA/isomiR here to study 3′ isomiRs. The mRNA segment of each chimeric read must exactly match a string in the mRNA sequences annotated in GENCODE[30]. Using this approach, we extracted all miRNA–mRNA and isomiR–mRNA interactions from each sample across all datasets.

The extracted IsomiRs in IsomiRdb

We downloaded the isomiR repertoire table from isomiRdb[11]. This table links each sequence to its corresponding mature miRNA and assigns an isoLabel indicating whether the sequence is "canonical" or variants. We classified sequences labeled as "canonical" in isomiRdb as miRNA sequences, whereas those with any other isoLabele as isomiRs. In total, this yielded 2573 annotated miRNAs and 3,274,688 annotated isomiRs (Fig. 2A). The number of isomiRs per miRNA ranged from 5 to 45,751, with a median of 258, a mean of 1233, and a standard deviation of 3549 (Fig. 2B). We next compared the extracted isomiRs from the six datasets with the annotated ones in isomiRdb to determine whether each predicted isomiR was already annotated, and to characterize the difference between annotated and unannotated miRNAs/isomiRs.

### Extracted miRNA–mRNA interactions in MiRTarBase and mirtargetline 2.0

We downloaded human miRNA–mRNA interaction data from miRTarBase[26] (hsa_MTI.csv file, version 10.0) and miRTargetLine 2.0[33]. miRTarBase included 2975 human miRNAs collectively targeting 16,976 mRNAs. Notably, some of these miRNAs, particularly from earlier annotation versions, may be inaccurate, as more recent annotations in miRbase (release 22.1) listed fewer miRNAs[31]. The number of mRNA targets per miRNA ranged from 1 to 3238, with a median of 354, a mean of 581, and a standard deviation of 579. miRTargetLink 2.0 included 203,478 experimentally validated miRNA–mRNA interactions, 6476 of which had strong experimental evidence. These miRTargetLink interactions involved 1184 miRNAs and 13,994 mRNAs.

We compared the extracted isomiR–mRNA and miRNA–mRNA interactions with those annotated in miRTarBase and mirTargetLink. For this comparison, we considered an isomiR–mRNA interaction as equivalent to a miRNA–mRNA interaction if the isomiR was a variant of the corresponding miRNA.

Properties of extracted isomiRs/miRNAs and their interactions with mRNAs

We examined the nucleotide composition of miRNAs with a high number of extracted isomiRs compared with that of miRNAs with fewer extracted isomiRs in our datasets. For each miRNA, we calculated the percentage of each nucleotide type. We then applied the Mann-Whitney test[34] to compare the two groups of percentages to see whether the percentages in one group were different from those in the other group. Similarly, we compared the nucleotide content of miRNAs with more extracted targets with that of miRNAs with fewer targets. Additionally, we investigated the nucleotide content of matching 6-mers between miRNAs that had matching seed regions and miRNAs that did not. A matching 6-mer was defined as a six-nucleotide RNA segment that was present in both an isomiR/miRNA and more than 90% of its mRNA targets. We considered a miRNA/isomiR having a matching seed region if it had a matching 6-mer started at positions 2 to 4 of this miRNA/isomiR sequence.

We also analyzed preferred mutation positions in miRNAs using a binomial test. A mutation position was defined as a nucleotide site within the mature miRNA where at least 5% of its isomiRs showed variation. For each miRNA with n isomiRs, we computed the p value for observing k or more isomiRs with variations at a specific position under the null hypothesis of random mutations across all 22 nucleotide positions in the mature miRNA. Positions with p values less than 0.05 were considered statistically significant, indicating a preferred mutation site at the corresponding position of the miRNA.

(A)   (B)
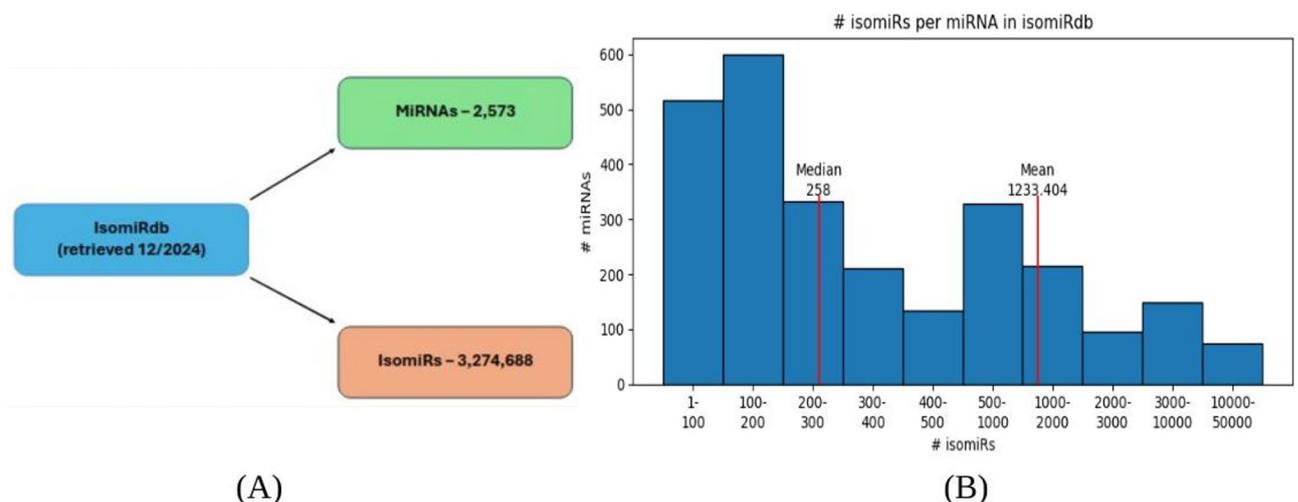
Fig. 2. **A** miRNAs and isomiRs in isomiRdb. **B** #isomiRs per miRNA in isomiRdb.

Results

Identification of over four hundred novel IsomiRs

We identified 1233 miRNAs interacting with 17,845 mRNAs across six datasets (Table 1). Among these, 407 miRNAs had at least one corresponding isomiR. Notably, 100 of these 407 miRNAs were represented exclusively by isomiRs, with no reference miRNA sequences detected. In total, we identified 1747 unique isomiRs that appeared at least 10 times across the datasets. The average, median, and standard deviation of the number of chimeric reads supporting each isomiR were 251, 37, and 2955, respectively. This distribution, with a low median and high standard deviation, suggested that while most isomiRs appeared only a few dozen times, a small subset was highly abundant. Specifically, 1298 isomiRs were supported by 10–99 reads, 395 by 100–999 reads, 49 by 1000–9999 reads, and five by more than 10,000 reads.

We studied the mutation positions in these 1747 isomiRs. We found that 536 (30.7%) had changes in their 5′. Interestingly, all these isomiRs also had changes in their 3′. In total, there were 1062 (60.8%) isomiRs having variations in their 3′ end. There might be even more isomiRs with variations in their 3′, since we copied the missed 3′ of miRNAs to miRNA segments in chimeric reads where the segments were shorter than the corresponding miRNAs.

We compared the isomiRs identified in our study with those curated in isomiRdb[11]. While isomiRdb included 3,274,688 isomiRs from 2573 miRNAs, our study identified 1747 isomiRs from 1233 miRNAs. Among these, 1277 (73.1%) were also found in isomiRdb, suggesting a high degree of reliability in our extracted isomiRs. In contrast, only 0.03% of isomiRs in isomiRdb were found in our study, reflecting the narrower scope and biological conditions of the six datasets we analyzed. To make a fairer comparison, we focused on the 1197 miRNAs common to both datasets. Within this subset, our 1747 isomiRs were compared with 2,604,710 isomiRs in isomiRdb. Of these, 1277 (73%) matched entries in isomiRdb, representing only 0.05% of the total isomiRs in isomiRdb.

The remaining 26.9% (470) of our extracted isomiRs not found in isomiRdb were still likely to be biologically valid. First, these isomiRs were observed multiple times across samples and datasets. On average, each unsupported isomiR appeared in 4.0 distinct samples and 1.4 dataset, compared with 3.4 samples and 1.3 datasets for supported isomiRs. Second, the read counts for these unsupported isomiRs were similar to those supported isomiRs. The median and mean number of chimeric reads for supported isomiRs were 38 and 254, respectively, compared to 33 and 245 for unsupported ones. These similarities in the number of samples, datasets and chimeric reads strongly suggest that the unsupported isomiRs are equally reliable and that existing isomiR annotations remain incomplete.

To further investigate the highly abundant isomiRs, we examined the 54 isomiRs supported by more than 1000 chimeric reads. To rule out experimental artifacts, we assessed their distribution across datasets. These isomiRs were present in an average of 23 distinct samples from multiple datasets, suggesting that their abundance reflects true biological activity rather than noise or dataset-specific artifacts. Additionally, we examined the chimeric read abundance of the corresponding reference miRNAs. These 54 isomiRs originated from 33 reference miRNAs. Interestingly, 28 of these miRNAs also produced at least one other isomiR supported by fewer than 100 reads, indicating differential expression of isomiR from the same reference miRNA. Of the remaining five miRNAs, four had only one detected isomiR, while the fifth, hsa-miR-30b-3p, had a second isomiR supported by over 200 reads. On average, these 33 miRNAs had a median of 13 and a mean of 18.7 associated isomiRs, highlighting the diversity and differential expression of isomiRs derived from a single miRNA.

miRNAs with higher adenine and thymine but lower cytosine content tend to have more isomiRs and mRNA targets

As mentioned above, we identified at least one isomiR for each of the 407 miRNAs, with 100 of these miRNAs represented solely by isomiRs. The number of isomiRs per miRNA ranged from 1 to 101, with a mean of 4 and a median of 1. Specifically, 227 miRNAs were associated with a single isomiR, 53 with two, 22 with three, 14 with four, and 91 with five or more isomiRs.

To understand why certain miRNAs give rise to more isomiRs, we examined the nucleotide composition of their reference sequences (Materials and Methods). Our analysis revealed that miRNAs with more isomiRs tend to have a higher content of adenine and thymine and a lower content of cytosine (Supplementary Table S3). For instance, when comparing miRNAs with 1–4 isomiRs with those with five or more, the differences in adenine, thymine, and cytosine content were statistically significant, with p values of 0.015, 0.009, and 0.010, respectively (Table 2).

Dataset	Sample information	miRNAs	isomiRs	mRNAs	miRNA–mRNA interactions	isomiR–mRNA interactions
CLASH	9 samples from Flp-In T-REx 293, female	405	111	9219	63,126	15,379
CLEAR-CLIP	12 samples from Huh-7.5, male	262	152	2549	17,617	5540
qCLASH	9 samples from TIVE-LTC, unknown gender	543	431	15,549	2,397,074	104,048
eCLIP	13 samples from HEK293T, female	812	1466	17,039	1,776,264	221,029
Kozer	3 samples from A375, female	511	286	14,305	601,513	63,280
Fields	21 samples from HCT116, male	525	173	11,278	262,192	29,739
Total	67 samples	1233	1747	17,845	5,117,786	439,015

Table 1. The number of miRNAs, isomiRs, and their interacting mRNAs across datasets.

Nucleotide type	miRNAs with 1 to 4 isomiRs vs those with >=5 isomiRs		miRNAs with >=1000 targets vs those with <1000 targets	
	p value	Corrected p value	p value	Corrected p value
A	0.015	0.06	4.62E−5	1.85E−4
C	0.010	0.04	0.042	0.168
G	0.805	1	0.477	1
T	0.009	0.036	0.026	0.094

Table 2. Nucleotide preference of MiRNAs with more IsomiRs and more targets.

We also examined whether specific nucleotide positions in miRNAs were more prone to mutations that generate isomiRs (Materials and Methods). Our analysis showed that the seed region was relatively conserved, with mutations occurring less frequently than in non-seed regions. In contrast, mutation rates were significantly higher outside the seed region, with 12 non-seed positions showing a p value < 0.05 and 10 positions with p value < 0.01. The mutation frequency exhibited a bimodal distribution, with notable peaks at positions 10 and 16 (Supplementary Table S4).

Next, we investigated the number of mRNA targets associated with each miRNA extracted from chimeric reads. Considering miRNAs and their associated isomiRs together, the number of mRNA targets per miRNA ranged from 1 to 12,910, with a mean of 697 and a standard deviation of 1843. The high standard deviation indicated considerable variation in the target counts across miRNAs. Specifically, 810 miRNAs had between 1 and 99 targets, 250 had 100 to 999 targets, 164 had 1000 to 9999 targets, and 9 miRNAs had over 10,000 targets.

To explore the factors distinguishing miRNAs with a large number of targets from those with fewer, we compared the nucleotide composition of 173 miRNAs with ≥ 1000 targets to that of the remaining 1060 miRNAs with fewer targets, using the Mann–Whitney test[34]. We found that miRNAs with many targets had significantly more adenines ($p < 4.62 \times 10^{-5}$) and thymines ($p < 0.05$), and significantly fewer cytosines ($p < 0.05$), confirming the nucleotide bias observed in isomiR-rich miRNAs (Table 2).

We also focused on the nine miRNAs with more than 10,000 mRNA targets. All but one of these miRNAs (hsa-miR-3168) belonged to miRNA families with at least two homologous members, and six had three or more homologs in miRBase[31] (Supplementary Table S5). Homologous miRNAs are those with identical mature sequences located in different genomic regions, such as hsa-let-7a-5p and hsa-let-7f-5p. Since only 268 of the 1233 miRNAs in our study have three or more homologs in miRBase, the probability of observing six of the nine high-target miRNAs with such homology is statistically significant ($p = 0.005$). This suggests that miRNA family expansion contributes to broader target coverage.

We also examined whether these nine miRNAs had more isomiRs than others. Remarkably, five of these miRNAs had between 30 and 60 isomiRs each. Since only seven of the 1233 miRNAs in our dataset had 30 or more isomiRs, the probability of five of them appearing among the nine is extremely low ($p = 1.1 \times 10^{-10}$). This result reinforces the idea that miRNA family size contributes to both isomiR diversity and target multiplicity. Although hsa-miR-3168 lacked homologs, it exhibited other characteristics that might explain its large target set: it was shorter (17 nucleotides compared to the typical 22) and had a higher adenine content (0.29 vs. an average of 0.22). These features may enhance its ability to bind a wider range of mRNA targets.

Finally, we studied the 100 miRNAs that were represented exclusively by their isomiRs in these datasets (Supplementary Table S6). We reviewed whether the original studies reported the absence of the reference miRNAs and the presence of their isomiRs. These studies rarely mentioned isomiRs or distinguished isomiRs from miRNAs, making it difficult to confirm our observations. We also analyzed the chimeric reads mapped to these miRNAs and their corresponding isomiRs. Notably, no chimeric reads were mapped to the miRNAs, while more than 100 chimeric reads were mapped to each isomiR on average. Moreover, these isomiRs appeared in more than 3.3 samples and more than 1.3 datasets on average, suggesting that their presence instead of their reference miRNAs is biologically meaningful, despite being unexplored in the literature.

IsomiRs significantly expand the target repertoire of their reference MiRNAs

Chimeric reads provide direct evidence of miRNA/isomiR–mRNA interactions. Across the six datasets analyzed, we inferred 830,336 miRNA–mRNA interactions and 132,530 isomiR–mRNA interactions. To assess the reliability and novelty of these interactions, we compared our findings with curated interactions in miRTarBase and miRTargetLink 2.0 (Materials and Methods).

We identified 1233 miRNAs targeting 17,845 mRNAs, of which 1184 miRNAs and 9451 mRNAs were also represented in miRTarBase. These overlapping entities accounted for 613,272 curated interactions in miRTarBase and 781,371 interactions in our datasets. Notably, 57,861 of the curated interactions were also detected in our study (precision = 0.074, recall = 0.033) (Table 3). The relatively low recall and precision values may stem from several factors, including the incompleteness of miRTarBase annotations, the mis-annotations (e.g., the 2975 miRNAs used), the limited sample diversity in our study, and the condition-specific nature of miRNA/isomiR–mRNA interactions. Despite this, interactions identified in both our study and miRTarBase appeared in an average of 4.2 samples, suggesting that these interactions are reproducible and biologically relevant. Interestingly, 14.8% of interactions not supported by miRTarBase were observed in more than 4.2 samples, implying that many of these unsupported interactions may be genuine but uncurated.

Dataset	#miRNAs	#mRNAs	#inferred interactions	#curated interactions	#common interactions	Precision	Recall
miRTarBase	1184	15,582	781,371	613,242	57,861	0.074	0.094
miRTargetLink all	1184	12,936	738,828	193,146	42,472	0.057	0.220
miRTargetLink strong	441	2351	141,378	6279	2837	0.020	0.452

Table 3. Comparison of extracted miRNA/isomiR–mRNA interactions with curated ones.

Similarly, we compared our miRNA/isomiR–mRNA interactions with those in miRTargetLink 2.0 (Table 3). For the common miRNAs and mRNAs in our study and miRTargetLink, we extracted 738,828 miRNA/isomiR–mRNA interactions and identified 42,472 interactions in miRTargetLink (precision = 0.057 and recall = 0.220). When focusing on the strongly validated interactions in miRTargetLink, we observed a precision of 0.020 and an improved recall 0.452. The improved recall indicated the good quality of the inferred interactions in our study, while the low precision suggested that a large percentage of miRNA/isomiR–mRNA interactions are still not identified and curated.

Next, we explored the independent contributions of isomiRs to miRNA-mRNA targeting. We focused on the 307 miRNAs for which both the reference miRNA and at least one isomiR were detected. We found that, on average, 18.9% of the targets for each miRNA-isomiR group were uniquely attributable to isomiRs, with a median of 2.7%. For instance, hsa-miR-30b-3p and its two isomiRs collectively targeted 629 mRNAs: 46 (7.3%) were unique to the reference miRNA, 539 (85.7%) were unique to the isomiRs, and 44 (7.0%) were shared. Strikingly, in 47 (15.3%) of the 307 cases, isomiRs contributed more unique targets than the corresponding reference miRNAs. These miRNAs, target mRNAs, and isomiRs did occur in samples, suggesting that the absence of reference miRNA binding in some cases was not due to the lack of expression of the miRNA or the mRNA. These analyses point to a biologically meaningful shift in targeting specificity, emphasizing the need to account for isomiRs in target prediction. We also studied the number of chimeric reads supporting individual interactions. On average, an miRNA–mRNA interaction was supported by 6.9 chimeric reads, whereas an isomiR–mRNA interaction had 3.3 chimeric reads. Interestingly, for 69 of the 307 miRNAs, isomiR–mRNA interactions were supported by more reads than miRNA–mRNA interactions (Supplementary Table S7).

To explore why certain targets were bound exclusively by isomiRs, we analyzed seed region sequences. Among the 307 miRNAs, 301 had isomiRs that bound targets not bound by the reference miRNA. These 301 miRNAs produced 1265 distinct isomiRs with unique targets. We found that 304 (24.0%) of these isomiRs, corresponding to 149 (49.5%) of the 301 miRNAs, had at least one nucleotide alteration in their seed regions. Surprisingly, the remaining 961 isomiRs (76.0%), associated with 152 miRNAs (50.5%), had unchanged seed sequences compared to their reference miRNAs. This suggests that in many cases, regions outside the seed may play a dominant role in isomiR-mediated targeting, as further explored in the next section.

A subset of MiRNAs and IsomiRs prefer non-seed binding to their targets

Previous studies have established that certain miRNAs can bind their targets through non-seed regions (positions other than 2–7)[20,21]. However, the prevalence and characteristics of non-seed binding remain unclear[11,14–16,21,22,35]. The collected chimeric reads provided a unique opportunity to systematically investigate the extent of non-seed binding of miRNAs and isomiRs. To explore this, we analyzed 1133 miRNAs and 1747 isomiRs identified across six datasets, excluding the 100 miRNAs represented exclusively by isomiRs. We further focused on a subset of 862 miRNAs and 1500 isomiRs, each with at least six mRNA targets. This cutoff of six ensured a false discovery rate of less than one false positive per miRNA/isomiR.

For each miRNA or isomiR, we performed local alignments between all possible 6-mers within the miRNA/isomiR sequence and the reverse complement of its corresponding mRNA targets. For a miRNA or isomiR, we retained only 6-mers that matched more than 90% of its associated target sequences. These frequently observed 6-mers were considered the common binding motifs of a given miRNA or isomiR[36]. Using this approach, we found that 544 miRNAs and 1405 isomiRs had at least one such common 6-mer motif. Among these, 435 miRNAs (80.0%) and 1290 isomiRs (91.8%) had their common 6-mers overlapping with the canonical seed region (defined as sharing at least three nucleotides with positions 2–7 of the mature miRNA sequence). Consequently, 20.0% of miRNAs and 8.2% of isomiRs appear to bind their targets without engaging the seed region. Furthermore, we found that seed-binding miRNAs and isomiRs frequently used additional non-seed regions for targeting. Specifically, 94.5% of seed-binding miRNAs and 95.7% of seed-binding isomiRs also had at least one non-seed 6-mer motif, suggesting that seed-based recognition is often complemented by non-seed interactions.

We then compared the characteristics of seed-binding and non-seed-binding miRNAs/isomiRs in four ways: (1) number of matching 6-mers: Seed-binding miRNAs had significantly more matching 6-mers (mean and median = 9) compared to non-seed-binding miRNAs (mean = 3, median = 2). The difference was highly significant (Mann–Whitney $p = 1.0E–37$); (2) location of matching regions: In non-seed-binding miRNAs/isomiRs, we observed that over half of the matching regions began between positions 6 and 8. Additionally, position 11 was part of the matching region in 73.5% of these miRNAs/isomiRs, highlighting the importance of the central region in non-seed-mediated targeting; (3) overall nucleotide composition: We found no significant difference in the overall nucleotide composition between seed-binding and non-seed-binding miRNAs/isomiRs; (4) nucleotide composition of matching regions: When analyzing the nucleotide content within the common 6-mers, we found that seed-binding motifs were significantly enriched in adenine and thymine and depleted in

Mutation rate per position
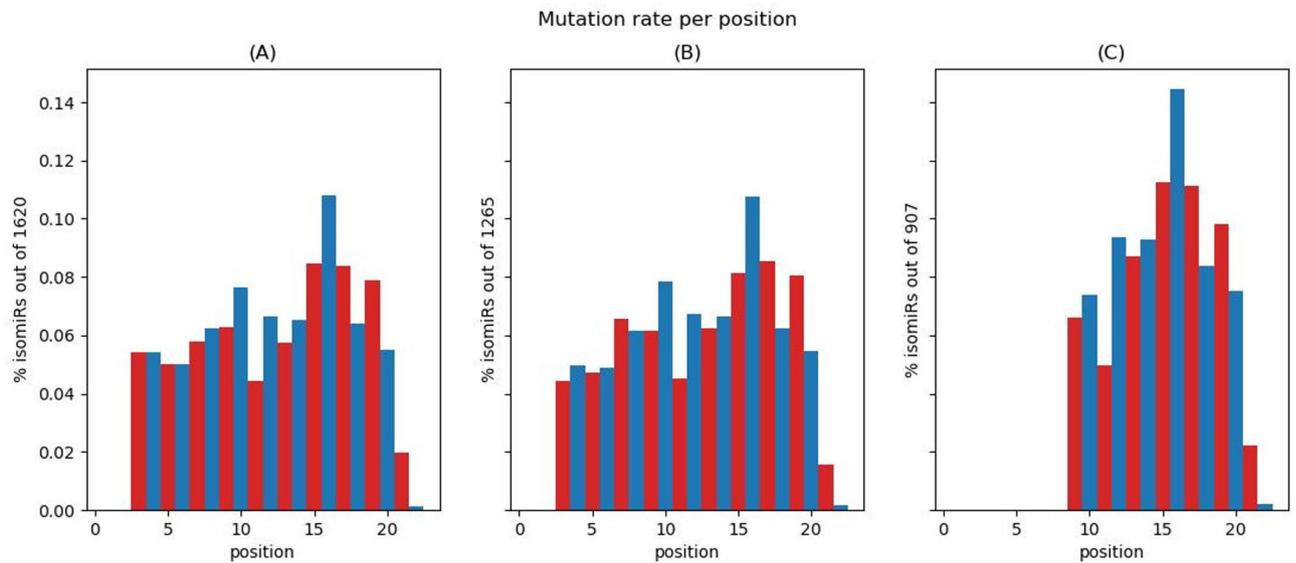
Fig. 3. The frequency of the mutated positions in isomiRs for **A** all 1620 isomiRs from 307 miRNAs; **B** all 1265 isomiRs in A with different targets from their miRNAs; **C** all 907 isomiRs in B with no seed-region mutations.

cytosine compared to non-seed-binding motifs (Mann–Whitney $p < 0.05$). This finding is consistent with earlier observations linking higher adenine/thymine content to increased numbers of isomiRs and mRNA targets.

Finally, we examined how mutations at specific positions of isomiRs influenced their targets. We compared three groups of isomiRs: (1) the 1620 isomiRs from the 307 miRNAs with the presence of both isomiRs and miRNAs; (2) the 1265 isomiRs in the first group that had different targets from their miRNAs; and (3) the 907 isomiRs in the second group that had no seed-region mutations. As noted earlier, isomiRs often exhibited variations out of the seed regions, with two peak regions around positions 10 and 16 (Fig. 3A). The distribution of the mutated positions was similar when comparing isomiRs in the first two groups (Fig. 3B). However, the peak at position 16 became particularly pronounced in the third group (Fig. 3C), suggesting that the region around position 16 plays an important role in determining miRNA/isomiR targeting specificity.

Taken together, these results indicate that while seed-based recognition is predominant, a substantial fraction of miRNAs and isomiRs utilize non-seed regions for target binding. The characteristics of these non-canonical interactions suggest potential mechanisms of regulatory flexibility and underscore the importance of considering full-length miRNA and isomiR sequences in target prediction models[22].

Discussion

In this study, we analyzed 67 human CLASH and CLASH-like samples and identified 1747 unique isomiRs and 439,015 isomiR–mRNA interactions. Approximately three-quarters of these isomiRs were previously annotated in isomiRdb[11], while only about 7.4% of the identified miRNA/isomiR–mRNA interactions were recorded in miRTarBase[26]. We found that miRNAs with higher adenine and thymine contents, but lower cytosine contents, tended to have more isomiRs and target more mRNAs. On average, 19.0%, and in some cases up to 98.1%, of a miRNA's targets were attributable exclusively to its isomiRs, even when these isomiRs shared identical seed sequences with their corresponding reference miRNAs.

Our results highlight significant gaps in the current understanding of isomiRs and their regulatory roles. Despite analyzing only 67 samples, 26.9% of the extracted isomiRs were absent from isomiRdb. These unannotated isomiRs occurred with similar frequencies and across comparable numbers of samples as curated isomiRs in isomiRdb, suggesting that existing isomiR annotations remain incomplete. Likewise, only a small fraction (6.7%) of the extracted miRNA/isomiR–mRNA interactions overlapped with miRTarBase records, and curated interactions were not better supported by chimeric read counts and sample occurrences than uncurated ones. These findings suggest that the diversity and regulatory impact of isomiR are likely underestimated.

Comparisons between our extracted miRNA/isomiR–mRNA interactions and the curated ones in miRTarBase and miRTargetLink further support this view. First, there is still a long way to go to curate miRNA–mRNA interactions. Among the three comparisons, the precision for comparison with larger curated datasets is larger, suggesting that the low precision is at least partially due to the limited number of curated interactions in current databases. Moreover, 14.8% of interactions not supported by miRTarBase appeared in as many samples as those supported interactions, implying that the curation of interactions is far from complete. Second, the number of samples and datasets in this study is limited. Although we recovered 45.2% strong interactions and 22.0% interactions in miRTargetLink, most known interactions are not observed in each of the three comparisons. With more samples and datasets in the future, the recall may be significantly improved. Third, the precision and recall in these comparisons are underestimated. We showed that the recall increased with the high confidence of the curated interactions, implying that the extracted interactions are of good quality. We also demonstrated that

the unsupported interactions appeared in as many samples and datasets as the curated ones. It is thus promising to focus on novel interactions we extracted and validate them experimentally in the future.

We found 100 miRNAs with only isomiRs in our datasets. We checked the number of chimeric reads supporting these isomiRs and miRNAs. We indeed found an average of 100 chimeric reads for each of these isomiRs but did not find any chimeric reads for these miRNAs (Supplementary Table S6). We also failed to find reports of this surprising phenomenon in literature. Given our earlier observation that isomiRs of the same miRNA can vary dramatically in abundance, these cases may reflect highly expressed isomiRs functionally replacing their reference miRNAs in certain context. This phenomenon warrants further investigation into the biological roles and evolution dynamics of miRNAs and isomiRs.

Although we attempted to include all available CLASH and CLASH-like datasets, our analysis was limited to 67 samples. As additional datasets become available, our findings can be revisited and validated on a larger scale. Moreover, current experimental methods often fail to capture the full-length sequences of miRNAs/isomiRs and their complete target mRNA fragments. Improved techniques that preserve these full-length chimeric segments will enable more precise characterization of miRNA/isomiR targeting. Finally, without functional assays or complementary datasets, we could not assess the downstream biological effects of the identified interactions. Future studies integrating transcriptomic, proteomic, or reporter assay data will b crucial for elucidating the functional significance of miRNA- and isomiR-mediated regulation.

Data availability

The CLASH, CLEAR-CLIP, qCLASH, chimeric eCLIP, and Fields et al. data were downloaded, respectively, from accession numbers GSE50452, GSE73059, GSE101978, GSE198251, and GSE164634. The Kozer et al. data is downloaded from (https://www.ebi.ac.uk/biostudies/arrayexpress/studies/E-MTAB-8591) . Supplementary Table S2 is available at (https://figshare.com/articles/dataset/Table_S1/28958909?file=54310421) . Supplementary Tables S1, S3 - S5 are available at (https://figshare.com/articles/dataset/Supplementary_tables_S2_-_S4/28959281?file=57067808) . Supplementary Table S6 is available at (https://figshare.com/articles/dataset/Table_S6/29876195) . Supplementary Table S7 is available at (https://figshare.com/articles/dataset/Table_S7/29876192) .

References

1. Morin, R. D. et al. Application of massively parallel sequencing to MicroRNA profiling and discovery in human embryonic stem cells. *Genome Res.* **18**, 610–621. https://doi.org/10.1101/gr.7179508 (2008).
2. Cloonan, N. et al. MicroRNAs and their IsomiRs function cooperatively to target common biological pathways. *Genome Biol.* **12** https://doi.org/10.1186/gb-2011-12-12-r126 (2011).
3. van der Kwast, R. V. C. T., Woudenberg, T., Quax, P. H. A. & Nossent, A. Y. MicroRNA-411 and its 5′-IsomiR have distinct targets and functions and are differentially regulated in the vasculature under ischemia. *Mol. Ther.* **28**, 157–170. https://doi.org/10.1016/j.ymthe.2019.10.002 (2020).
4. Tomasello, L., Distefano, R., Nigita, G. & Croce, C. M. The MicroRNA family gets wider: the IsomiRs classification and role. *Front. Cell. Dev. Biol.* **9**, 668648. https://doi.org/10.3389/fcell.2021.668648 (2021).
5. Wu, C. W. et al. A comprehensive approach to sequence-oriented IsomiR annotation (CASMIR): demonstration with IsomiR profiling in colorectal neoplasia. *BMC Genom.* **19**, 401. https://doi.org/10.1186/s12864-018-4794-7 (2018).
6. Mercey, O. et al. Characterizing IsomiR variants within the microRNA-34/449 family. *Febs Lett.* **591**, 693–705. https://doi.org/10.1002/1873-3468.12595 (2017).
7. Zelli, V. et al. Emerging role of IsomiRs in cancer: state of the Art and recent advances. *Genes-Basel* **12** https://doi.org/10.3390/genes12091447 (2021).
8. van der Kwast, R. V. C. T., Quax, P. H. A. & Nossent, A. Y. An emerging role for IsomiRs and the MicroRNA epitranscriptome in neovascularization. *Cells-Basel* **9** https://doi.org/10.3390/cells9010061 (2020).
9. Wagner, V., Meese, E. & Keller, A. The intricacies of isomirs: from classification to clinical relevance. *Trends Genet.* **40**, 784–796. https://doi.org/10.1016/j.tig.2024.05.007 (2024).
10. Wang, Y., Goodison, S., Li, X. M. & Hu, H. Y. Prognostic cancer gene signatures share common regulatory motifs. *Sci. Rep.* **7** https://doi.org/10.1038/s41598-017-05035-3 (2017).
11. Aparicio-Puerta, E. et al. IsomiRdb: MicroRNA expression at isoform resolution. *Nucleic Acids Res.* **51** (D179-D185). https://doi.org/10.1093/nar/gkac884 (2023).
12. Ros, B. D. Tumor IsomiR encyclopedia (TIE): a pan-cancer database of MiRNA isoforms. *Bioinformatics* **37**, 3023–3025. https://doi.org/10.1093/bioinformatics/btab172 (2021).
13. Zhang, Y. W. et al. IsomiR bank: a research resource for tracking IsomiRs. *Bioinformatics* **32**, 2069–2071. https://doi.org/10.1093/bioinformatics/btw070 (2016).
14. Nersisyan, S. et al. isomiRTar: a comprehensive portal of pan-cancer 5′-isomiR targeting. *Peerj* **10** https://doi.org/10.7717/peerj.14205 (2022).
15. Chen, Y. H. & Wang, X. W. MiRDB: an online database for prediction of functional MicroRNA targets. *Nucleic Acids Res.* **48**, D127–D131. https://doi.org/10.1093/nar/gkz757 (2020).
16. McGeary, S. E. et al. The biochemical basis of MicroRNA targeting efficacy. *Science* **366**, 1470–. https://doi.org/10.1126/science.aav1741 (2019).
17. Talukder, A., Zhang, W. C., Li, X. M. & Hu, H. Y. A deep learning method for miRNA/isomiR target detection. *Sci. Rep.* **12** https://doi.org/10.1038/s41598-022-14890-8 (2022).
18. Helwak, A., Kudla, G., Dudnakova, T. & Tollervey, D. Mapping the human MiRNA interactome by CLASH reveals frequent noncanonical binding. *Cell* **153**, 654–665. https://doi.org/10.1016/j.cell.2013.03.043 (2013).
19. Ding, J., Li, X. M. & Hu, H. Y. MicroRNA modules prefer to bind weak and unconventional target sites. *Bioinformatics* **31**, 1366–1374. https://doi.org/10.1093/bioinformatics/btu833 (2015).
20. Ding, J., Li, X. M. & Hu, H. Y. TarPmiR: a new approach for MicroRNA target site prediction. *Bioinformatics* **32**, 2768–2775. https://doi.org/10.1093/bioinformatics/btw318 (2016).
21. Ding, J., Li, X. M. & Hu, H. Y. CCmiR: a computational approach for competitive and cooperative MicroRNA binding prediction. *Bioinformatics* **34**, 198–206. https://doi.org/10.1093/bioinformatics/btx606 (2018).

22. Talukder, A., Li, X. M. & Hu, H. Y. Position-wise binding preference is important for MiRNA target site prediction. *Bioinformatics* **36**, 3680–3686. https://doi.org/10.1093/bioinformatics/btaa195 (2020).
23. Gay, L. A., Turner, P. C. & Renne, R. Modified Cross-Linking, Ligation, and sequencing of hybrids (qCLASH) to identify MicroRNA targets. *Curr. Protoc.* **1** https://doi.org/10.1002/cpz1.257 (2021).
24. Manakov, S. et al. Scalable and deep profiling of mRNA targets for individual MicroRNAs with chimeric eCLIP. (2022).
25. Moore, M. J. et al. MiRNA-target chimeras reveal miRNA 3′-end pairing as a major determinant of argonaute target specificity. *Nat. Commun.* **6** https://doi.org/10.1038/ncomms9864 (2015).
26. Cui, S. D. et al. MiRTarBase 2025: updates to the collection of experimentally validated microRNA-target interactions. *Nucleic Acids Res.* **53**, D147–D156. https://doi.org/10.1093/nar/gkae1072 (2024).
27. Kozar, I. et al. Cross-Linking ligation and sequencing of hybrids (qCLASH) reveals an unpredicted MiRNA targetome in melanoma cells. *Cancers* **13** https://doi.org/10.3390/cancers13051096 (2021).
28. Fields, C. J. et al. Sequencing of Argonaute-bound microRNA/mRNA hybrids reveals regulation of the unfolded protein response by microRNA-320a. *PLoS Genet.* **17** https://doi.org/10.1371/journal.pgen.1009934 (2021).
29. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for illumina sequence data. *Bioinformatics* **30**, 2114–2120. https://doi.org/10.1093/bioinformatics/btu170 (2014).
30. Harrow, J. et al. GENCODE: the reference human genome annotation for the ENCODE project. *Genome Res.* **22**, 1760–1774. https://doi.org/10.1101/gr.135350.111 (2012).
31. Kozomara, A., Birgaoanu, M. & Griffiths-Jones, S. MiRBase: from MicroRNA sequences to function. *Nucleic Acids Res.* **47**, D155–D162. https://doi.org/10.1093/nar/gky1141 (2019).
32. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2 (1990).
33. Kern, F. et al. MiRTargetLink 2.0-interactive MiRNA target gene and target pathway networks. *Nucleic Acids Res.* **49**, W409–W416. https://doi.org/10.1093/nar/gkab297 (2021).
34. Mann, H. B. & Whitney, D. R. On a test of whether one of two random variables is stochastically larger than the other. *Ann. Math. Stat.* **18**, 50–60. https://doi.org/10.1214/aoms/1177730491 (1947).
35. Athaya, T., Li, X. M. & Hu, H. Y. A deep learning method to integrate extracelluar MiRNA with mRNA for cancer studies. *Bioinformatics* **40** https://doi.org/10.1093/bioinformatics/btae653 (2024).
36. Ding, J., Cai, X., Wang, Y., Hu, H. & Li, X. ChIPModule: systematic discovery of transcription factors and their cofactors from ChIP-seq data. *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing*, 320–331 (2013).

Acknowledgements

Author contributions

H.H. and X.L. conceived and designed the study. M.W. performed the experiments. M.W., R.C.R., X.L., and H.H. analyzed the data. M.W., R.C.R., X.L., and H.H. wrote the manuscript. All authors proofread the manuscript.

Declarations

Competing interests

We declare that there is no conflict of interest regarding the publication of this article.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-21561-x.

Correspondence and requests for materials should be addressed to X.L. or H.H.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.