# JC

RK

2013-04-30


Cell

# Mapping the Human miRNA Interactome by CLASH Reveals Frequent Noncanonical Binding

Aleksandra Helwak,[1,2] Grzegorz Kudla,[1,2,3] Tatiana Dudnakova,[1] and David Tollervey[1,*]
[1]Wellcome Trust Centre for Cell Biology, The University of Edinburgh, Edinburgh EH9 3JR, UK
[2]These authors contributed equally to this work
[3]Present address: MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, The University of Edinburgh, Edinburgh EH4 2XU, UK
*Correspondence: d.tollervey@ed.ac.uk
http://dx.doi.org/10.1016/j.cell.2013.03.043

## SUMMARY

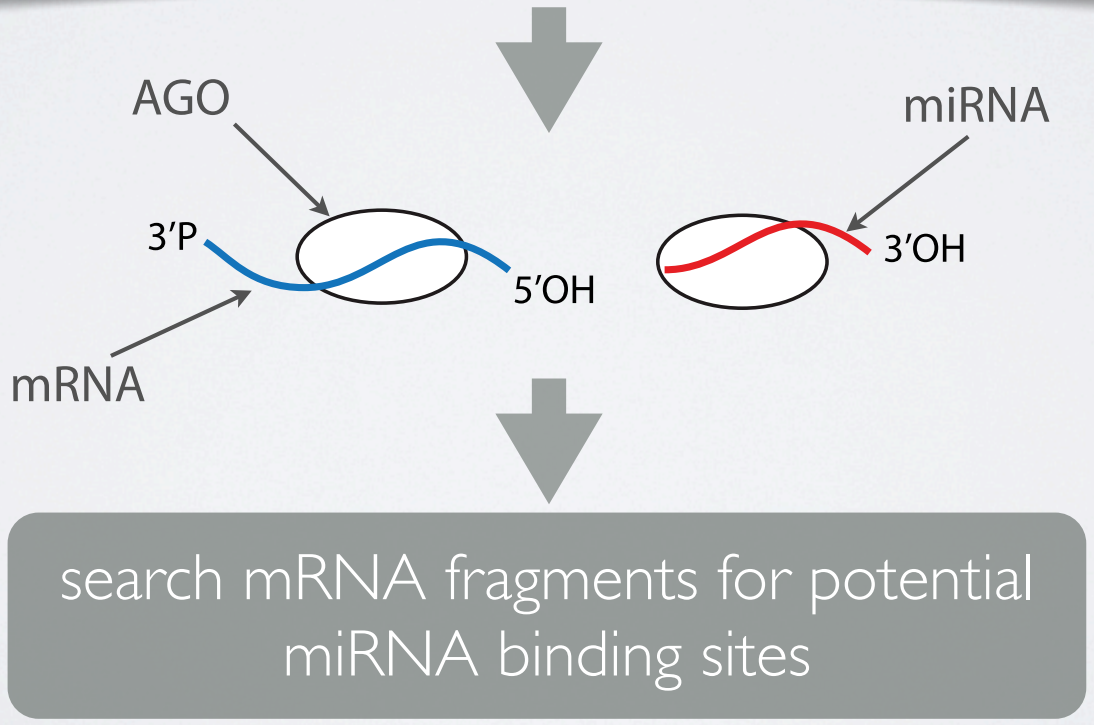
MicroRNAs (miRNAs) play key roles in gene regulation, but reliable bioinformatic or experimental identification of their targets remains difficult. To provide an unbiased view of human miRNA targets, we devel-

pairs with the target; (2) nucleotides paired outside the seed region stabilize interactions but are reported not to influence miRNA efficacy (Garcia et al., 2011; Grimson et al., 2007); and (3) functional miRNA targets are localized close to the extremes of the 3′ UTRs of protein-coding genes in relatively unstructured regions (Grimson et al., 2007). Recently,

# Argonaute HITS-CLIP decodes microRNA–mRNA interaction maps

Sung Wook Chi[1], Julie B. Zang[1], Aldo Mele[1] & Robert B. Darnell[1]

**MicroRNAs (miRNAs) have critical roles in the regulation of gene expression; however, as miRNA activity requires base pairing with only 6–8 nucleotides of messenger RNA, predicting target mRNAs is a major challenge. Recently, high-throughput sequencing of RNAs isolated by crosslinking immunoprecipitation (HITS-CLIP) has identified functional protein–RNA interaction sites. Here we use HITS-CLIP to covalently crosslink native argonaute (Ago, also called Eif2c) protein–RNA complexes in mouse brain. This produced two simultaneous data sets—Ago–miRNA and Ago–mRNA binding sites—that were combined with bioinformatic analysis to identify interaction sites between miRNA and target mRNA. We validated genome-wide interaction maps for miR-124, and generated additional maps for the 20 most abundant miRNAs present in P13 mouse brain. Ago HITS-CLIP provides a general platform for exploring the specificity and range of miRNA action *in vivo*, and identifies precise sequences for targeting clinically relevant miRNA–mRNA interactions.**




**HITS-CLIP**

Cell

# Mapping the Human miRNA Interactome by CLASH Reveals Frequent Noncanonical Binding

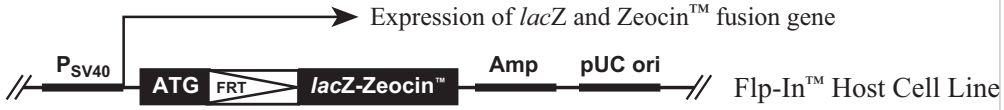
Aleksandra Helwak,[1,2] Grzegorz Kudla,[1,2,3] Tatiana Dudnakova,[1] and David Tollervey[1,*]
[1]Wellcome Trust Centre for Cell Biology, The University of Edinburgh, Edinburgh EH9 3JR, UK
[2]These authors contributed equally to this work
[3]Present address: MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, The University of Edinburgh, Edinburgh EH4 2XU, UK
*Correspondence: d.tollervey@ed.ac.uk
http://dx.doi.org/10.1016/j.cell.2013.03.043

## SUMMARY

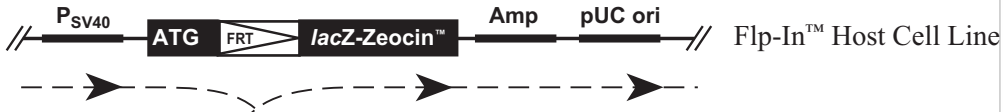
MicroRNAs (miRNAs) play key roles in gene regulation, but reliable bioinformatic or experimental identification of their targets remains difficult. To provide an unbiased view of human miRNA targets, we devel-

pairs with the target; (2) nucleotides paired outside the seed region stabilize interactions but are reported not to influence miRNA efficacy (Garcia et al., 2011; Grimson et al., 2007); and (3) functional miRNA targets are localized close to the extremes of the 3′ UTRs of protein-coding genes in relatively unstructured regions (Grimson et al., 2007). Recently,

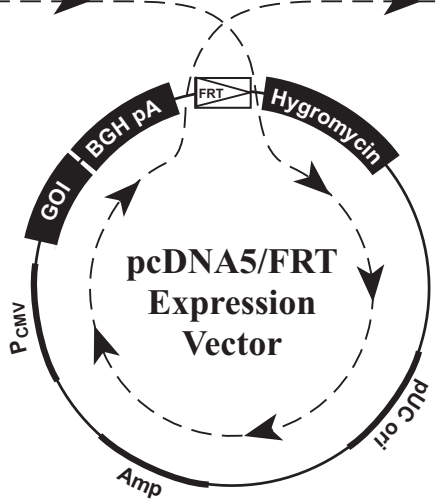
1. pFRT/*lac*Zeo is stably transfected into the mammalian cells of interest to generate the Zeocin™-resistant Flp-In™ Host Cell Line(s)



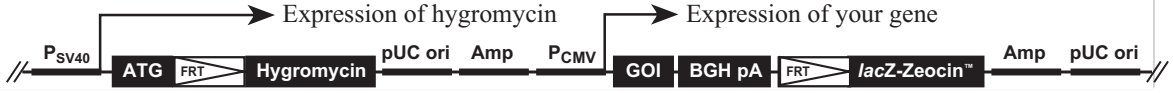

2. The pcDNA5/FRT expression vector containing your gene of interest (GOI) is cotransfected with pOG44 into the Flp-In™ Host Cell Line.


3. The Flp recombinase expressed from pOG44 catalyzes a homologous recombination event between the FRT sites in the host cells and the pcDNA5/FRT expression vector.




4. Integration of the expression construct allows transcription of the gene of interest (GOI) and confers hygromycin resistance and Zeocin™ sensitivity to the cells.



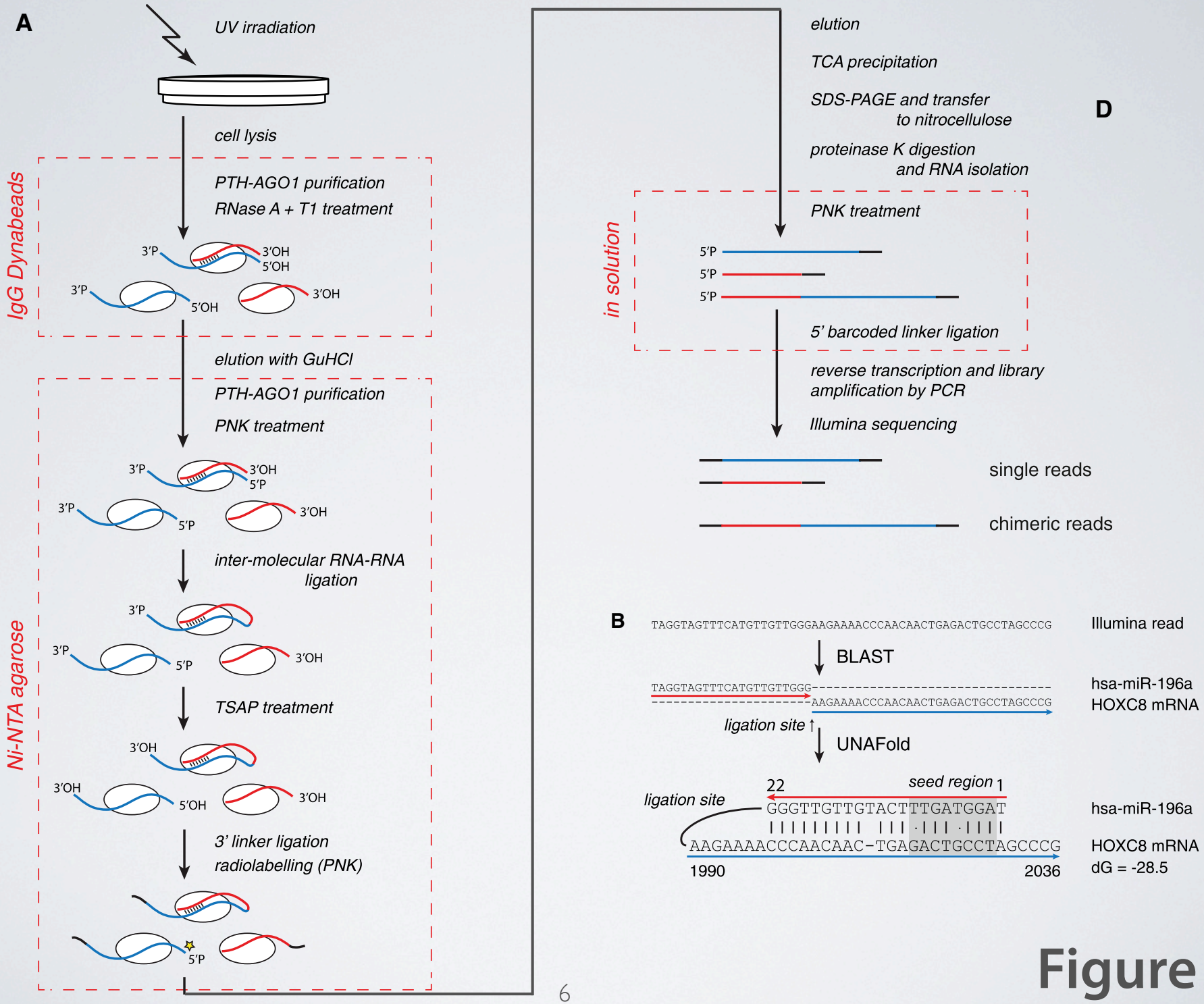
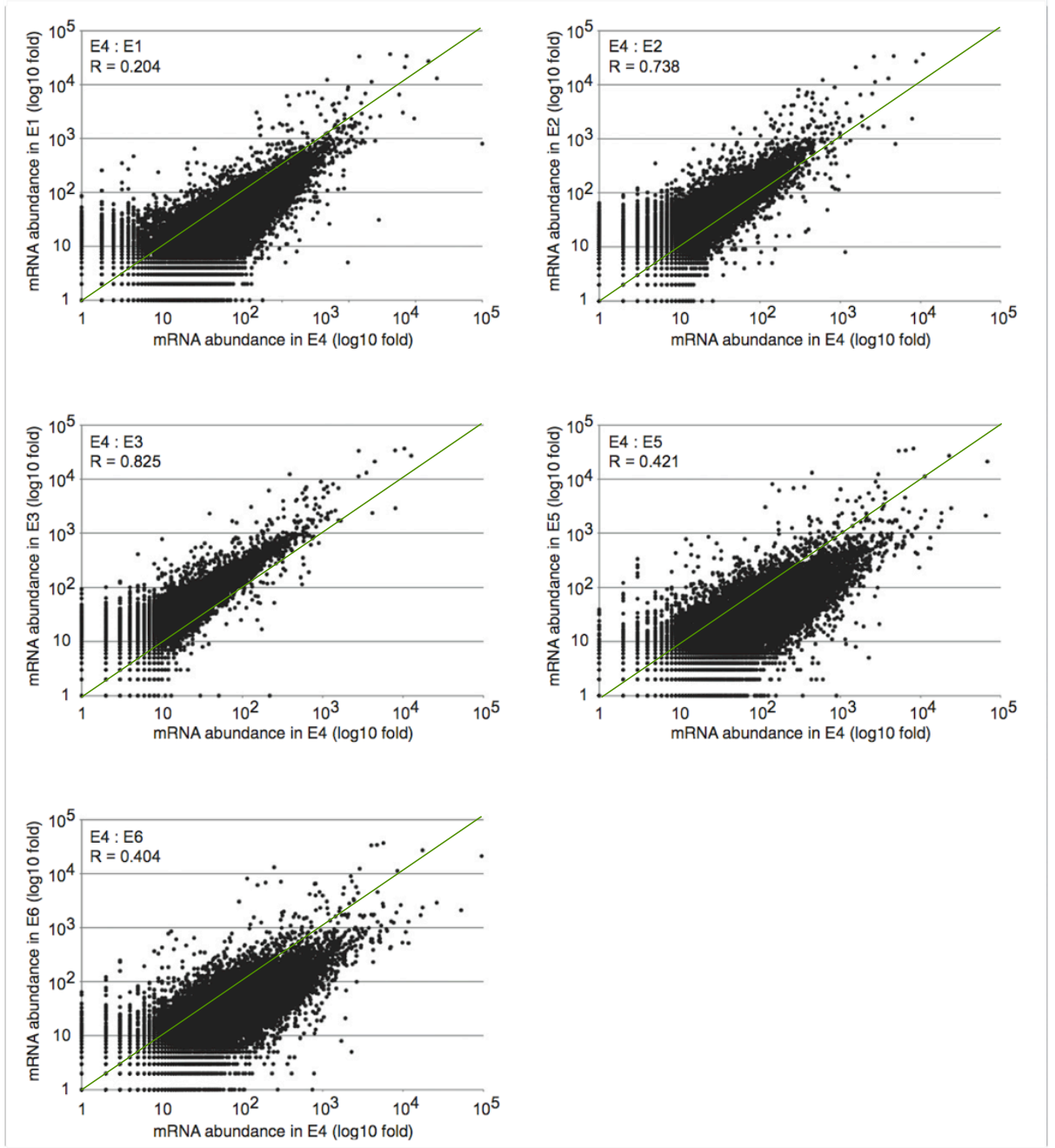

A
UV irradiation
cell lysis
IgG Dynabeads
PTH-AGO1 purification
RNase A + T1 treatment
3'P
3'OH
5'OH
elution with GuHCl
PTH-AGO1 purification
PNK treatment
5'P
inter-molecular RNA-RNA ligation
Ni-NTA agarose
TSAP treatment
3' linker ligation
radiolabelling (PNK)
elution
TCA precipitation
SDS-PAGE and transfer to nitrocellulose
proteinase K digestion and RNA isolation
in solution
PNK treatment
5' barcoded linker ligation
reverse transcription and library amplification by PCR
Illumina sequencing
D
single reads
chimeric reads
B
TAGGTAGTTTCATGTTGTTGGGAAGAAAACCCAACAACTGAGACTGCCTAGCCCG
Illumina read
BLAST
TAGGTAGTTTCATGTTGTTGGG
AAGAAAACCCAACAACTGAGACTGCCTAGCCCG
hsa-miR-196a
HOXC8 mRNA
ligation site
UNAFold
22
seed region
1
ligation site
GGGTTGTTGTACTTTGATGGAT
hsa-miR-196a
AAGAAAACCCAACAAC-TGAGACTGCCTAGCCCG
HOXC8 mRNA
1990
2036
dG = -28.5

**Figure 1**

enzymatic reactions at 37°C, not 20°C

different ligase for 3' linker

linker ligation performed in solution for 16 hours




3' linker ligation performed on beads, not gel & different ligase for 3' linker

linker ligation performed in solution for 1 hour

**Supp Fig 1**

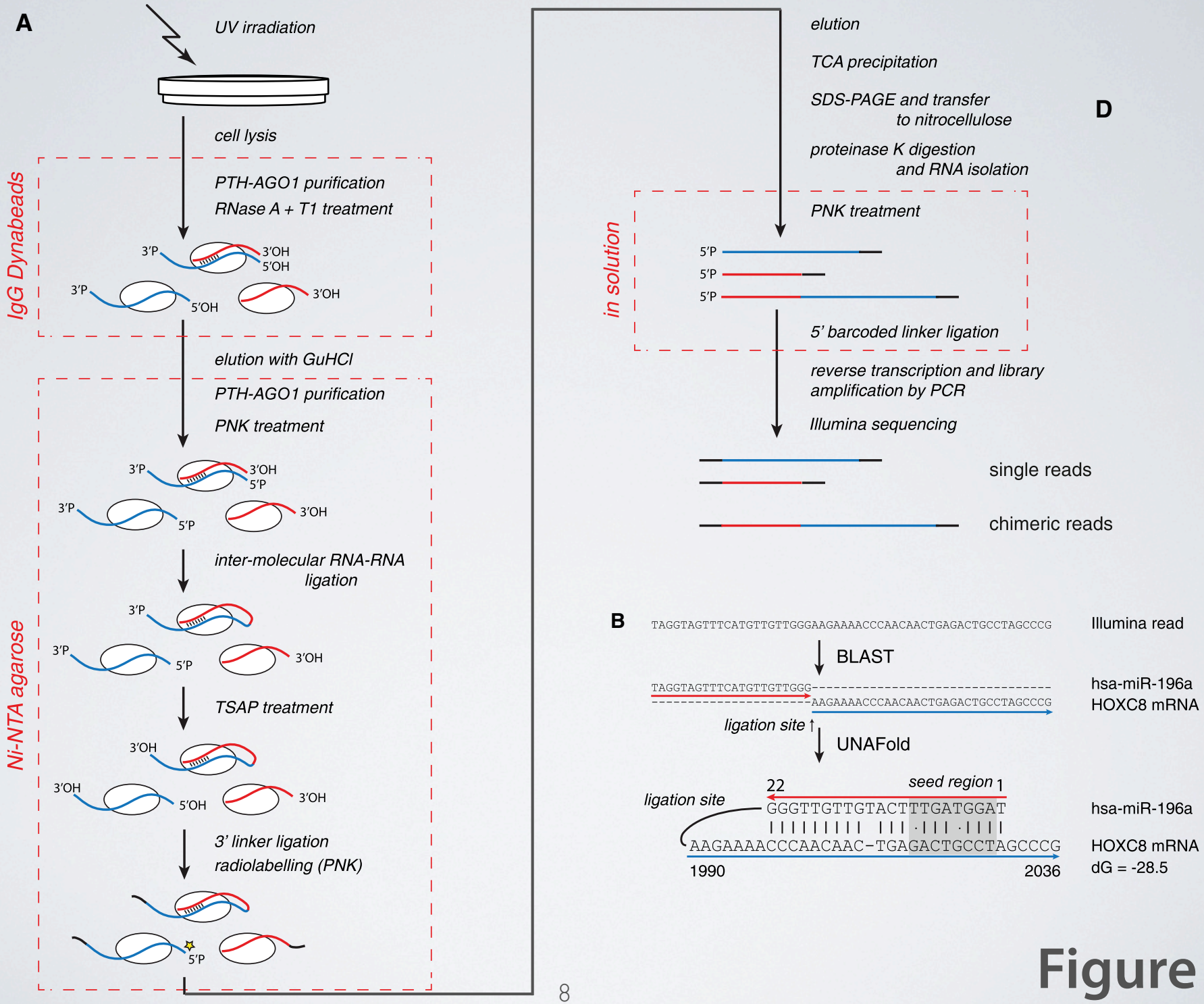
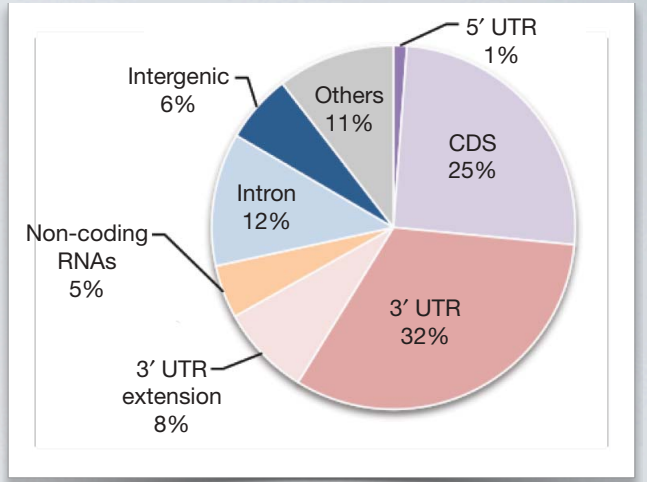
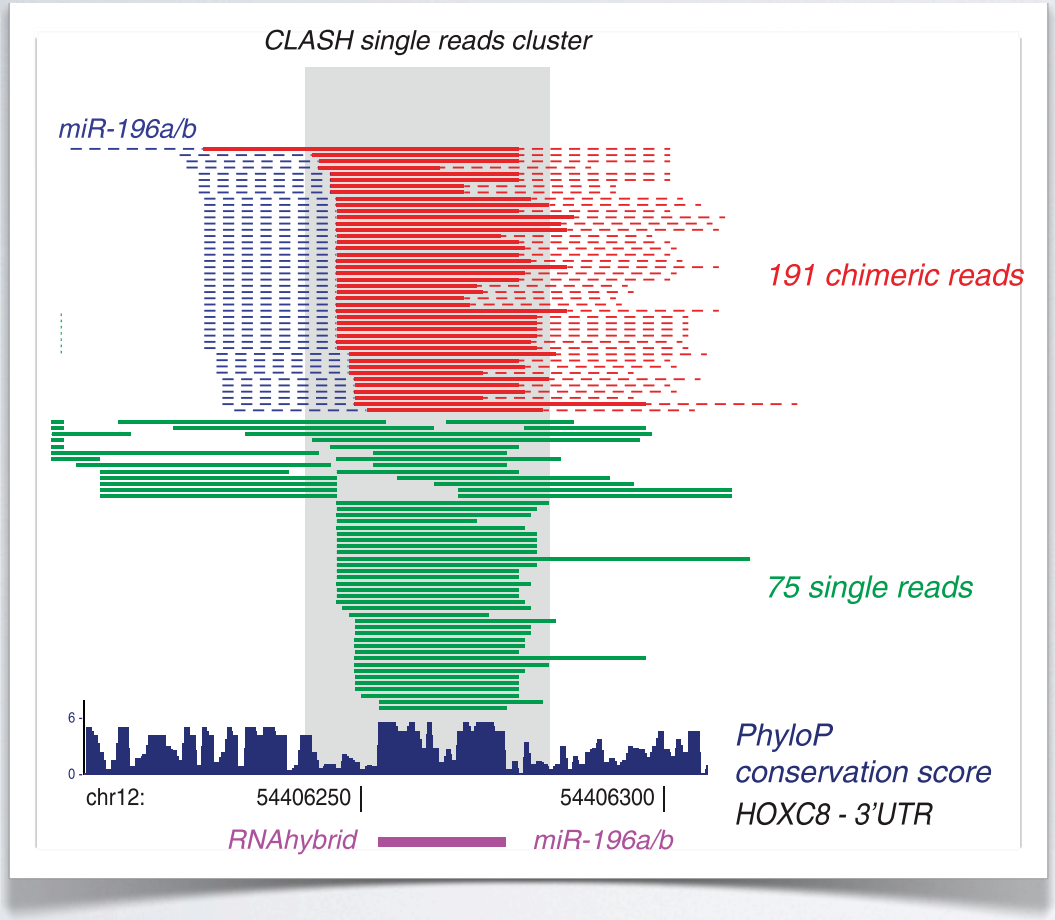

A
UV irradiation
cell lysis
IgG Dynabeads
PTH-AGO1 purification
RNase A + T1 treatment
3′P
3′OH
5′OH
elution with GuHCl
PTH-AGO1 purification
PNK treatment
5′P
Ni-NTA agarose
inter-molecular RNA-RNA ligation
TSAP treatment
3′ linker ligation
radiolabelling (PNK)
elution
TCA precipitation
SDS-PAGE and transfer to nitrocellulose
proteinase K digestion and RNA isolation
in solution
PNK treatment
5′ barcoded linker ligation
reverse transcription and library amplification by PCR
Illumina sequencing
single reads
chimeric reads
D
B
TAGGTAGTTTCATGTTGTTGGGAAGAAAACCCAACAACTGAGACTGCCTAGCCCG
Illumina read
BLAST
TAGGTAGTTTCATGTTGTTGGG
AAGAAAACCCAACAACTGAGACTGCCTAGCCCG
hsa-miR-196a
HOXC8 mRNA
ligation site
UNAFold
22
seed region
1
ligation site
GGGTTGTTGTACTTTGATGGAT
hsa-miR-196a
AAGAAAACCCAACAAC-TGAGACTGCCTAGCCCG
HOXC8 mRNA
dG = -28.5
1990
2036

**Figure 1**

# HITS-CLIP →




interactions
(chimeric reads) →

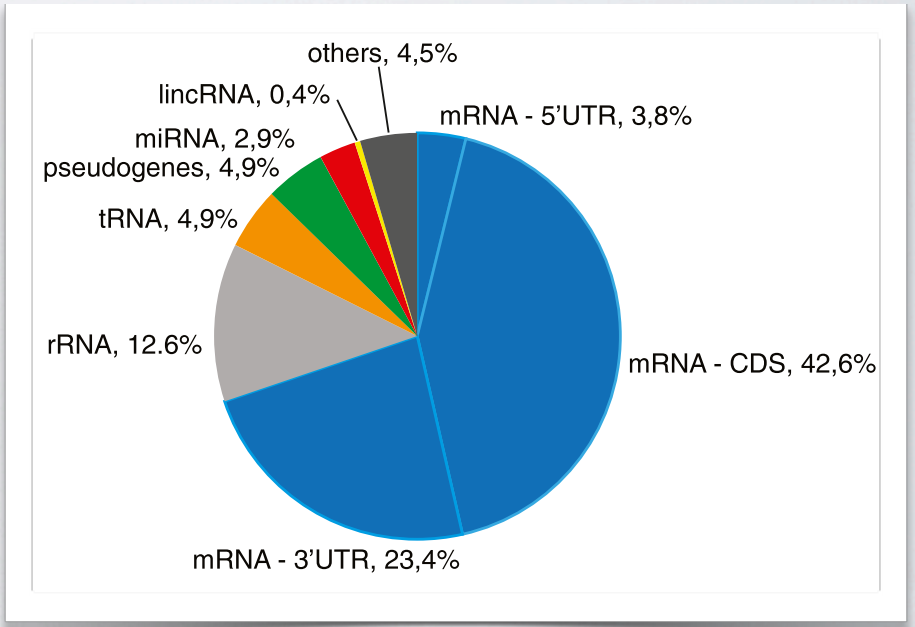



**Figure 1**

## Table S2A

| | ALL DATASETS | | DATASET E4 | |
|---|---|---|---|---|
| | Single reads | % | Single reads | % |
| **microRNA*** | 21,832,771 | 20.6 | 1,722,779 | 7.6 |
| **mRNA** | 20,134,598 | 19.0 | 3,767,064 | 16.5 |
| **tRNA** | 40,910,266 | 38.5 | 11,596,776 | 50.9 |
| **rRNA** | 13,491,752 | 12.7 | 3,709,881 | 16.3 |
| **snRNA** | 1,546,938 | 1.5 | 300,695 | 1.3 |
| **pseudogenes** | 957,987 | 0.9 | 183,904 | 0.8 |
| **lincRNA** | 519,538 | 0.5 | 138,261 | 0.6 |
| **snoRNA** | 489,682 | 0.5 | 88,446 | 0.4 |
| **other** | 6,288,134 | 5.9 | 1,272,717 | 5.6 |
| **Total:** | 106,171,666 | | 22,780,523 | |

## Table S2B

| | All | containing miRNAs | miRNA-mRNA (% of all) |
|---|---|---|---|
| **ALL DATASETS** | | | |
| CHIMERIC READS: | 1,700,221 | 563,984 | 66,304 (3.9%) |
| collapsed reads | 622,721 | 106,765 | 28,255 (4.5%) |
| number of supported interactions | 240,135 | 26,543 | 18,514 (7.7%) |
| number of identified miRNAs | | | 399 |
| number of identified mRNAs [ENSG/ENST] | | | 6,959 / 7,390 |
| **PUBLISHED DETAILED PROTOCOL (E4)** | | | |
| CHIMERIC READS: | 794,729 | 303,154 | 29,943 (3.8%) |
| collapsed reads | 290,221 | 58,337 | 17,237 (5.9%) |
| number of supported interactions | 110,148 | 16,386 | 12,198 (11.1%) |
| number of identified miRNAs | | | 340 |
| number of identified mRNAs [ENSG/ENST] | | | 5,427 / 5,705 |



**Supp Table 2**

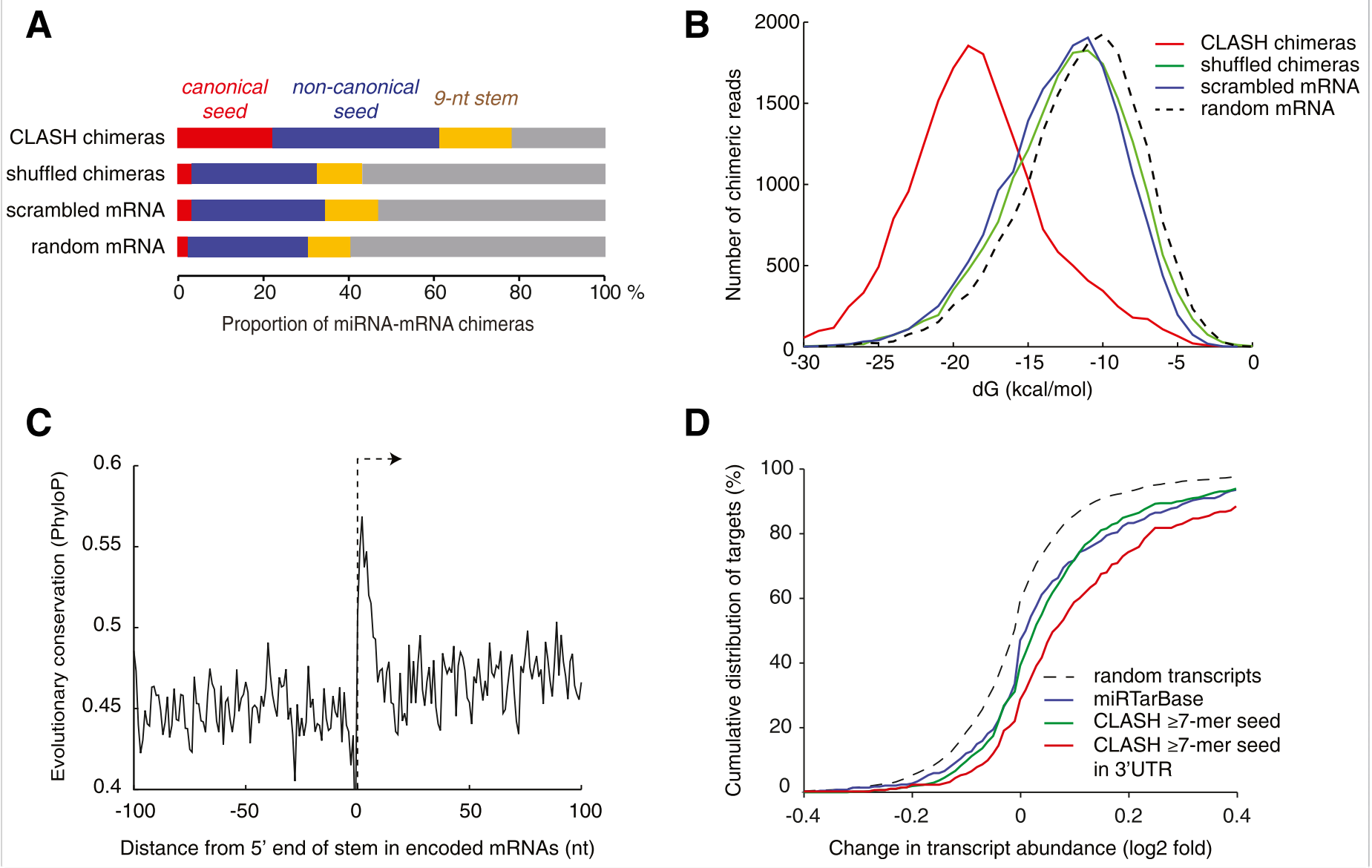
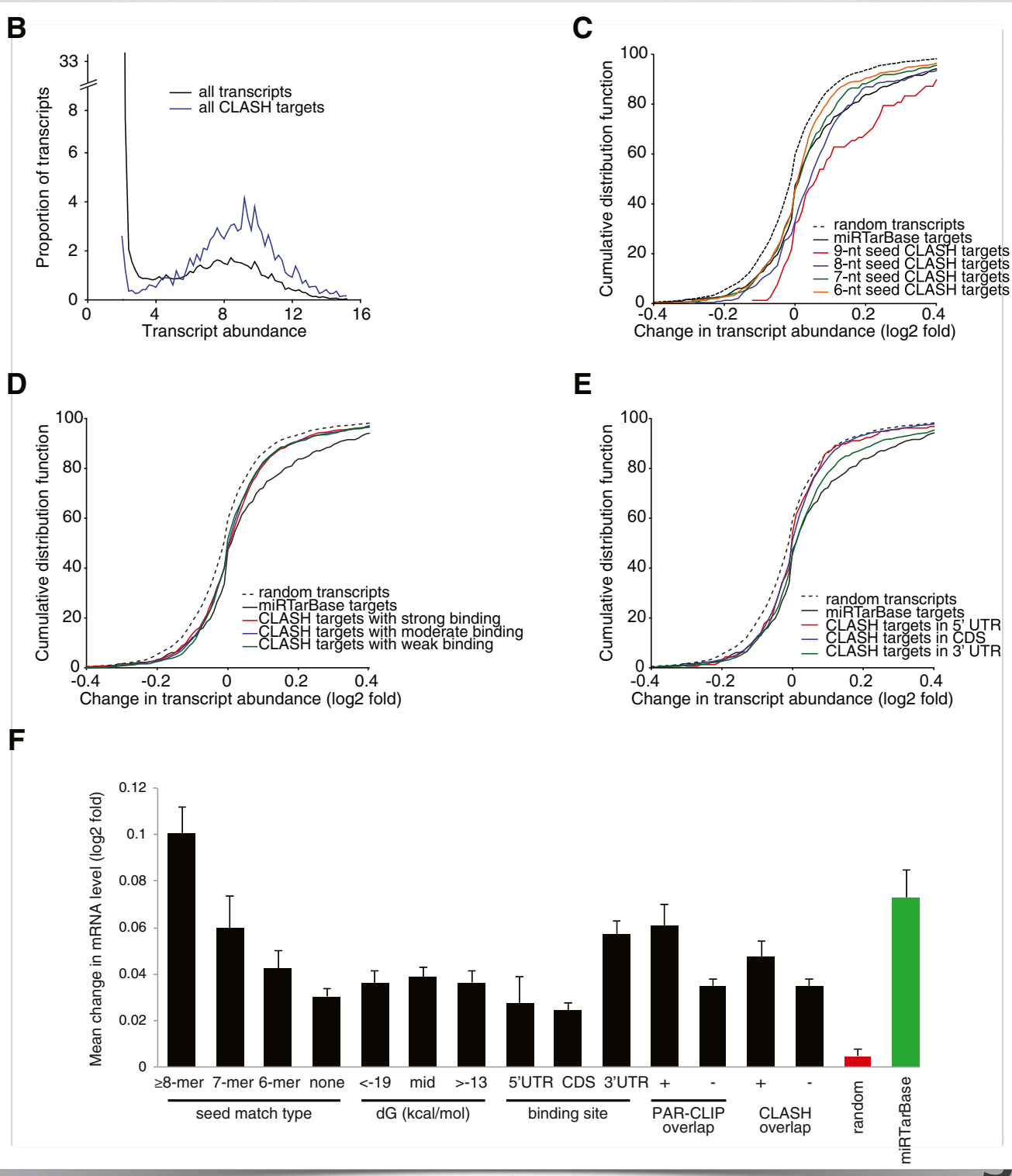

A
canonical seed
non-canonical seed
9-nt stem
CLASH chimeras
shuffled chimeras
scrambled mRNA
random mRNA
0
20
40
60
80
100 %
Proportion of miRNA-mRNA chimeras
B
Number of chimeric reads
2000
1500
1000
500
0
-30
-25
-20
-15
-10
-5
0
dG (kcal/mol)
CLASH chimeras
shuffled chimeras
scrambled mRNA
random mRNA
C
Evolutionary conservation (PhyloP)
0.6
0.55
0.5
0.45
0.4
-100
-50
0
50
100
Distance from 5' end of stem in encoded mRNAs (nt)
D
Cumulative distribution of targets (%)
100
80
60
40
20
0
-0.4
-0.2
0
0.2
0.4
Change in transcript abundance (log2 fold)
random transcripts
miRTarBase
CLASH ≥7-mer seed
CLASH ≥7-mer seed in 3'UTR

Figure 2


B
Proportion of transcripts
Transcript abundance
all transcripts
all CLASH targets
C
Cumulative distribution function
Change in transcript abundance (log2 fold)
random transcripts
miRTarBase targets
9-nt seed CLASH targets
8-nt seed CLASH targets
7-nt seed CLASH targets
6-nt seed CLASH targets
D
Cumulative distribution function
Change in transcript abundance (log2 fold)
random transcripts
miRTarBase targets
CLASH targets with strong binding
CLASH targets with moderate binding
CLASH targets with weak binding
E
Cumulative distribution function
Change in transcript abundance (log2 fold)
random transcripts
miRTarBase targets
CLASH targets in 5' UTR
CLASH targets in CDS
CLASH targets in 3' UTR
F
Mean change in mRNA level (log2 fold)
≥8-mer
7-mer
6-mer
none
<-19
mid
>-13
5'UTR
CDS
3'UTR
+
-
+
-
random
miRTarBase
seed match type
dG (kcal/mol)
binding site
PAR-CLIP overlap
CLASH overlap

Supp Fig 2

Table S4A

| | | PAR-CLIP clusters | CLASH single read clusters |
|---|---|---|---|
| | ***Number of interactions*** | *15,823* | *28,515* |
| **CLASH chimeras** | *18,514* | 1,596 (9%) | 3,066 (17%) |
| random localization within the same gene | *18,514* | 1,125 | 2,036 |
| random localization within a random gene with similar expression level | *18,514* | 540 | 968 |
| random localization within a random gene | *18,514* | 268 | 439 |
| **CLASH single read clusters** | *28,515* | 2,528 (9%) | |
| random localization within the same gene | *28,515* | 1,417 | |
| random localization within a random gene with similar expression level | *28,515* | 702 | |
| random localization within a random gene | *28,515* | 343 | |

Table S4B

| | | matches with CLASH | matches with control | enrichment |
|---|---|---|---|---|
| | ***Number of interactions*** | *6,248* | *6,248* | |
| **miRanda** | *687,208* | 411 | 29 | 14.2 × |
| **PicTar** | *205,263* | 224 | 9 | 24.9 × |
| **PITA** | *192,255* | 195 | 2 | 97.5 × |
| **RNAhybrid** | *992,584* | 310 | 25 | 12.4 × |
| **TargetScan** | *54,199* | 170 | 5 | 34.0 × |
| **all predictions** | *2,131,509* | 802 | 59 | 13.6 × |



**Supp Table 4**

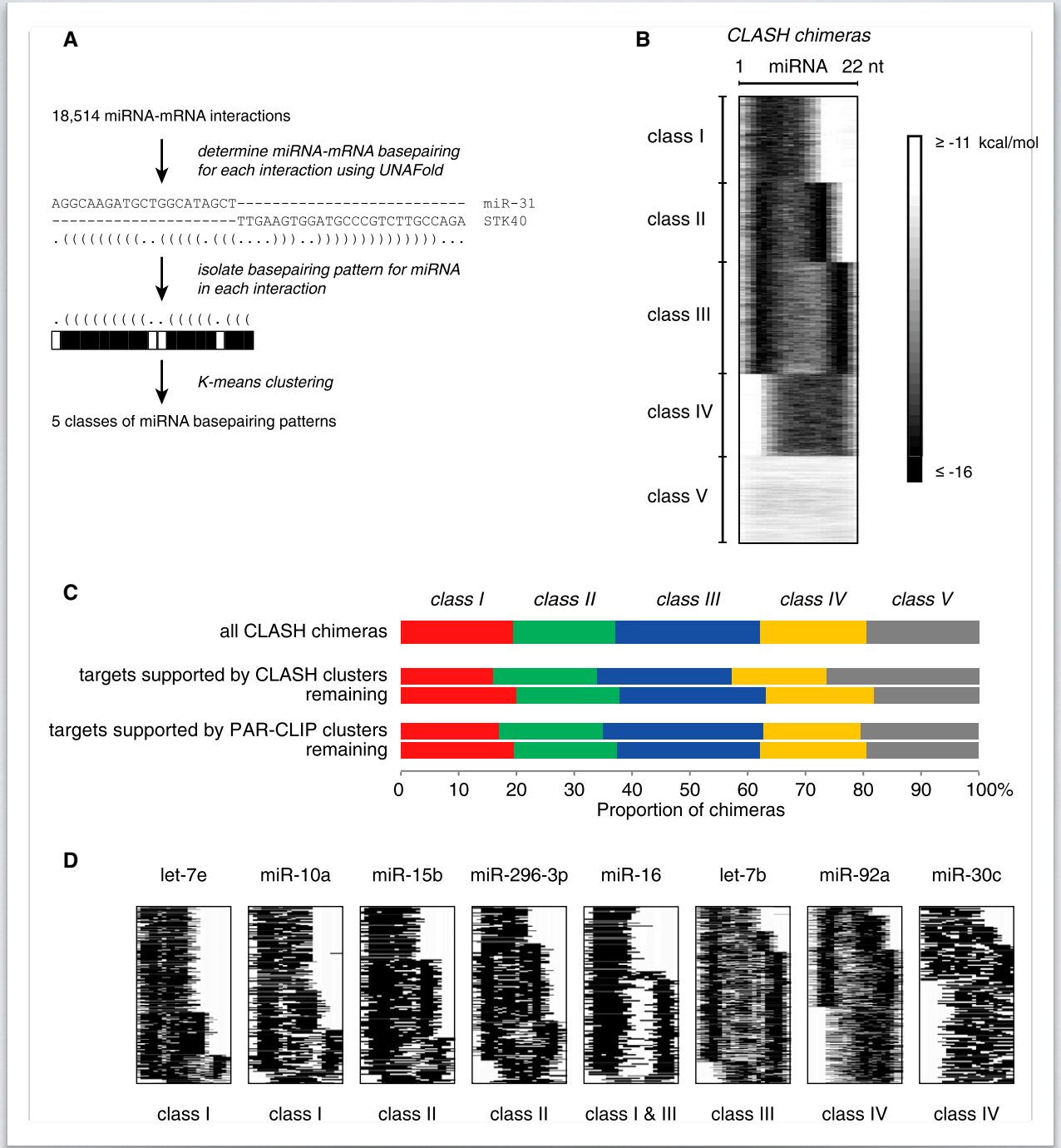

A
18,514 miRNA-mRNA interactions
determine miRNA-mRNA basepairing for each interaction using UNAFold
AGGCAAGATGCTGGCATAGCT---------- miR-31
----------TTGAAGTGGATGCCCGTCTTGCCAGA STK40
.((((((((((..((((((.(((....))))..))))))))))))))...
isolate basepairing pattern for miRNA in each interaction
.((((((((((..((((((.(((
K-means clustering
5 classes of miRNA basepairing patterns
B
CLASH chimeras
1 miRNA 22 nt
class I
class II
class III
class IV
class V
≥ -11 kcal/mol
≤ -16
C
class I
class II
class III
class IV
class V
all CLASH chimeras
targets supported by CLASH clusters remaining
targets supported by PAR-CLIP clusters remaining
0 10 20 30 40 50 60 70 80 90 100%
Proportion of chimeras
D
let-7e
miR-10a
miR-15b
miR-296-3p
miR-16
let-7b
miR-92a
miR-30c
class I
class I
class II
class II
class I & III
class III
class IV
class IV



Figure 3

**Table 1. Analysis of the Five miRNA-mRNA Base-Pairing Classes**

| Class | I | II | III | IV | V |
|---|---|---|---|---|---|
| Number of interactions | 3,594 | 3,293 | 4,630 | 3,389 | 3,608 |
| Number of base-paired nucleotides | 13.0 ± 0.04 | 15.3 ± 0.03 | 16.8 ± 0.03 | 14.6 ± 0.04 | 11.9 ± 0.05 |
| Number of base-paired nucleotides in seed | 5.2 ± 0.02 | 5.1 ± 0.02 | 5.0 ± 0.01 | 2.6 ± 0.02 | 3.3 ± 0.03 |
| Interaction energy (dG) | −18.3 ± 0.04 | −20.2 ± 0.06 | −20.5 ± 0.05 | −19.0 ± 0.05 | −11.1 ± 0.05 |
| PhyloP conservation score | 0.092 ± 0.017 | 0.127 ± 0.018 | 0.097 ± 0.017 | 0.011 ± 0.017 | 0.086 ± 0.018 |
| Efficiency of inhibition by miRNA | 0.042 ± 0.007 | 0.052 ± 0.009 | 0.047 ± 0.005 | 0.024 ± 0.005 | 0.039 ± 0.004 |
| Targets in 5′ UTR | 4.8% | 4.2% | 4.1% | 5.8% | 4.7% |
| Targets in CDS | 60.7% | 61.1% | 61.4% | 63.9% | 53.4% |
| Targets in 3′ UTR | 32.7% | 32.2% | 32.1% | 28.1% | 39.5% |

Table 1

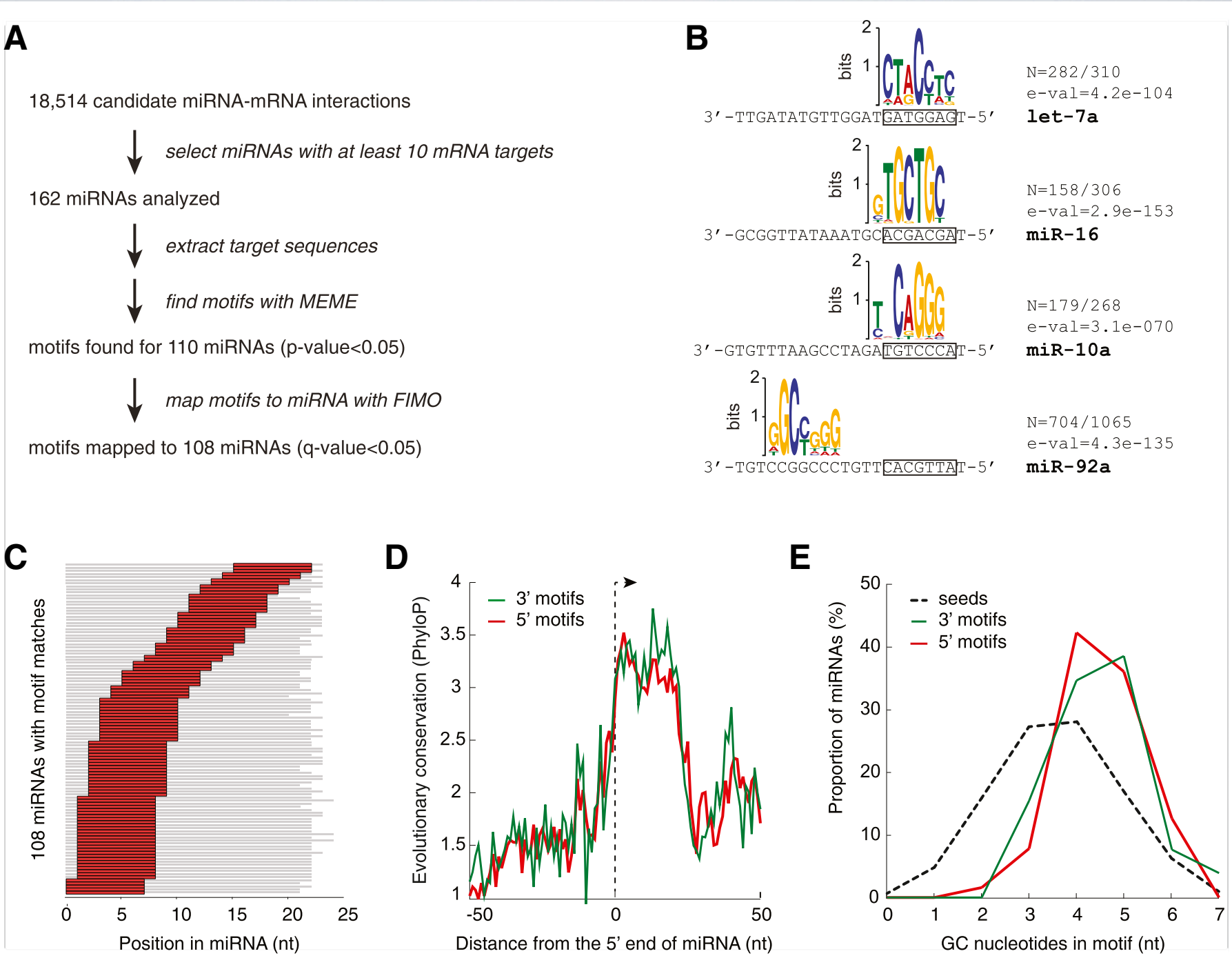

A
18,514 candidate miRNA-mRNA interactions
select miRNAs with at least 10 mRNA targets
162 miRNAs analyzed
extract target sequences
find motifs with MEME
motifs found for 110 miRNAs (p-value<0.05)
map motifs to miRNA with FIMO
motifs mapped to 108 miRNAs (q-value<0.05)
B
bits
3′-TTGATATGTTGGATGATGGAGT-5′
N=282/310
e-val=4.2e-104
let-7a
3′-GCGGTTATAAATGCACGACGAT-5′
N=158/306
e-val=2.9e-153
miR-16
3′-GTGTTTAAGCCTAGATGTCCCAT-5′
N=179/268
e-val=3.1e-070
miR-10a
3′-TGTCCGGCCCTGTTCACGTTAT-5′
N=704/1065
e-val=4.3e-135
miR-92a
C
108 miRNAs with motif matches
Position in miRNA (nt)
D
3′ motifs
5′ motifs
Evolutionary conservation (PhyloP)
Distance from the 5′ end of miRNA (nt)
E
seeds
3′ motifs
5′ motifs
Proportion of miRNAs (%)
GC nucleotides in motif (nt)

Figure 4

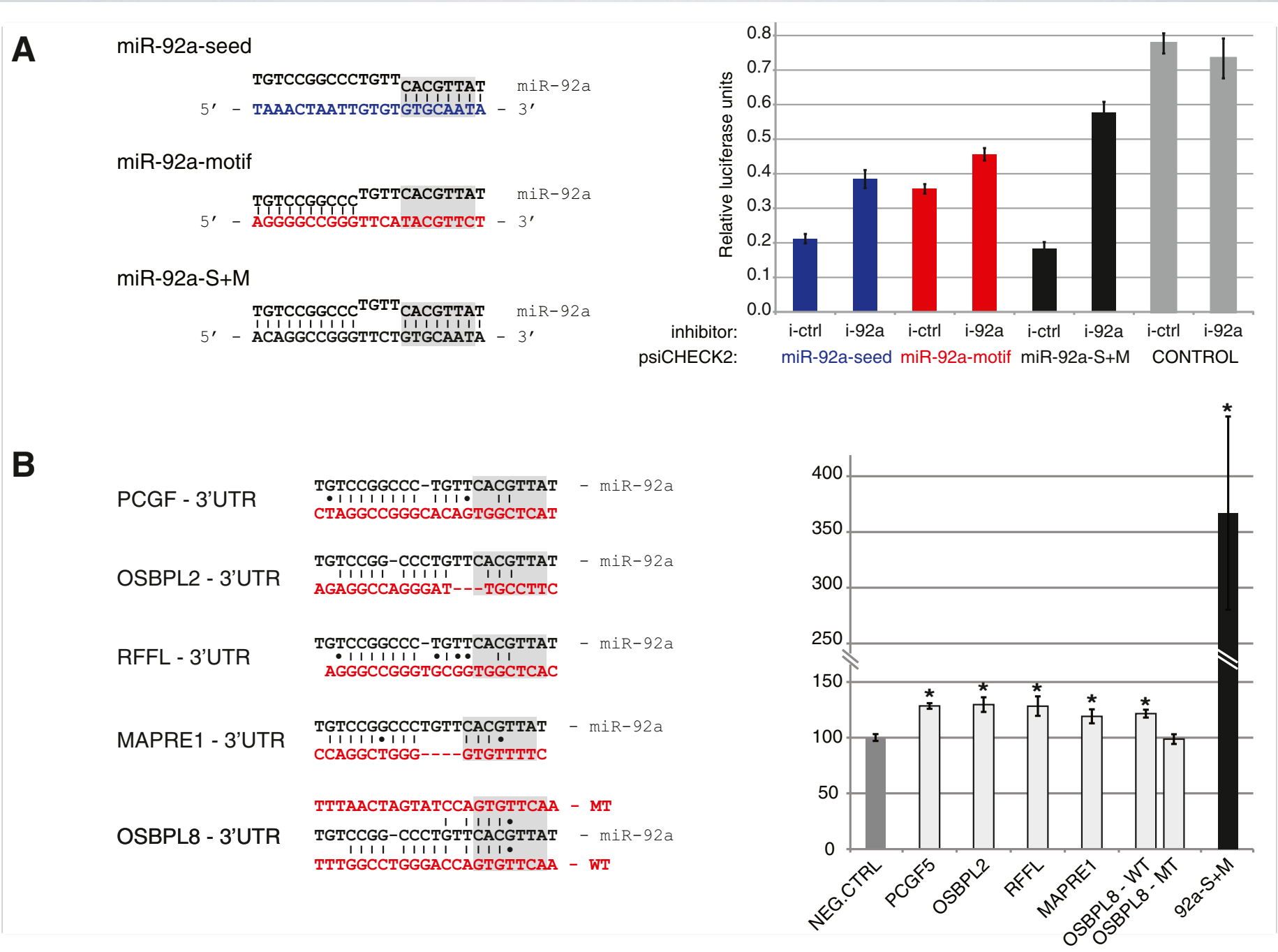

A
miR-92a-seed
TGTCCGGCCCTGTT CACGTTAT miR-92a
5′ - TAAACTAATTGTGTGTGCAATA - 3′
miR-92a-motif
TGTCCGGCCC TGTTCACGTTAT miR-92a
5′ - AGGGGCCGGGTTCATACGTTCT - 3′
miR-92a-S+M
TGTCCGGCCC TGTT CACGTTAT miR-92a
5′ - ACAGGCCGGGTTCTGTGCAATA - 3′
Relative luciferase units
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
inhibitor: i-ctrl i-92a i-ctrl i-92a i-ctrl i-92a i-ctrl i-92a
psiCHECK2: miR-92a-seed miR-92a-motif miR-92a-S+M CONTROL
B
PCGF - 3'UTR
TGTCCGGCCC-TGTTCACGTTAT - miR-92a
CTAGGCCGGGCACAGTGGCTCAT
OSBPL2 - 3'UTR
TGTCCGG-CCCTGTTCACGTTAT - miR-92a
AGAGGCCAGGGAT---TGCCTTC
RFFL - 3'UTR
TGTCCGGCCC-TGTTCACGTTAT - miR-92a
AGGGCCGGGTGCGGTGGCTCAC
MAPRE1 - 3'UTR
TGTCCGGCCCTGTTCACGTTAT - miR-92a
CCAGGCTGGG----GTGTTTTC
OSBPL8 - 3'UTR
TTTAACTAGTATCCAGTGTTCAA - MT
TGTCCGG-CCCTGTTCACGTTAT - miR-92a
TTTGGCCTGGGACCAGTGTTCAA - WT
400
350
300
250
150
100
50
0
NEG.CTRL
PCGF5
OSBPL2
RFFL
MAPRE1
OSBPL8 - WT
OSBPL8 - MT
92a-S+M

Figure 5

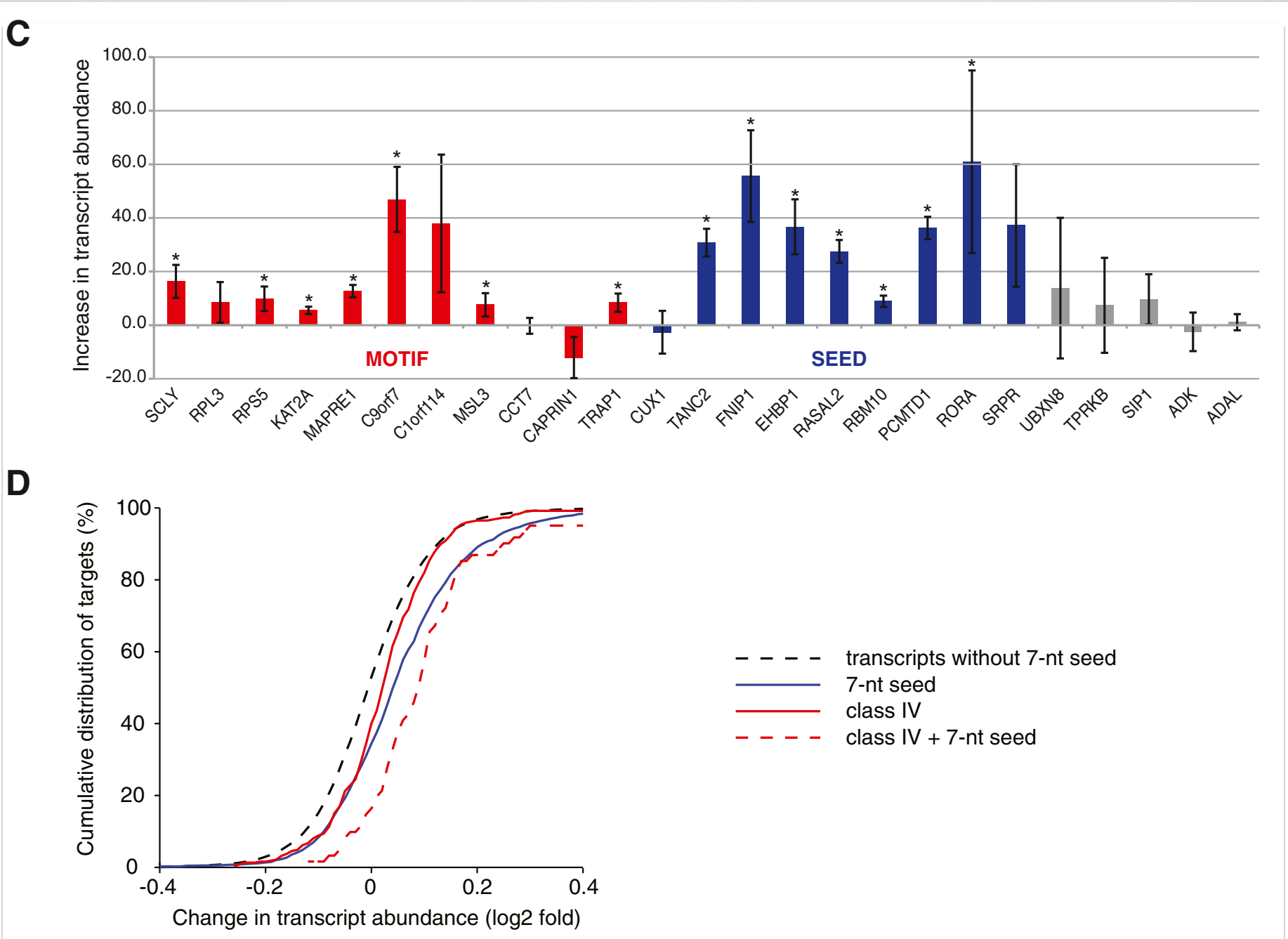
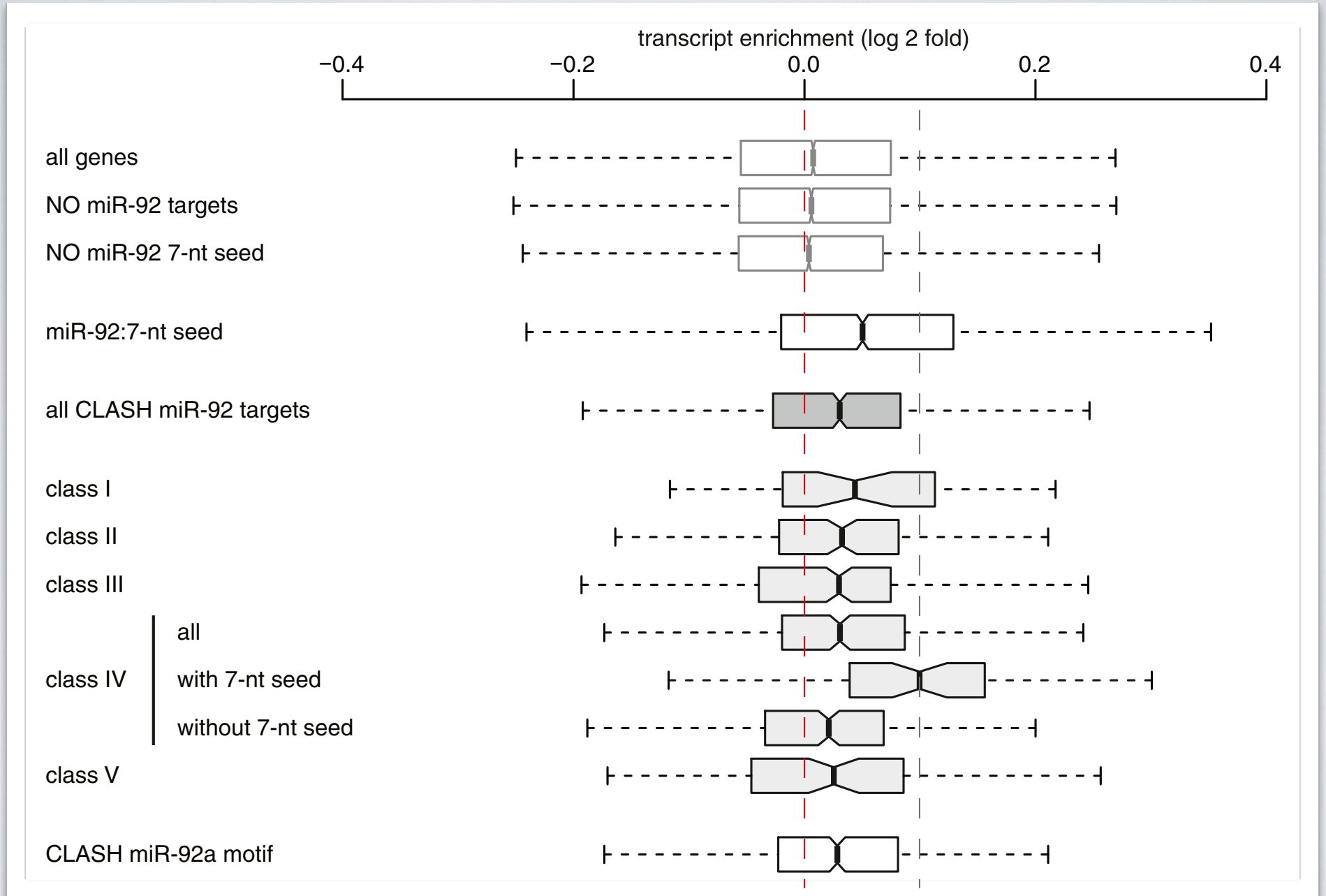
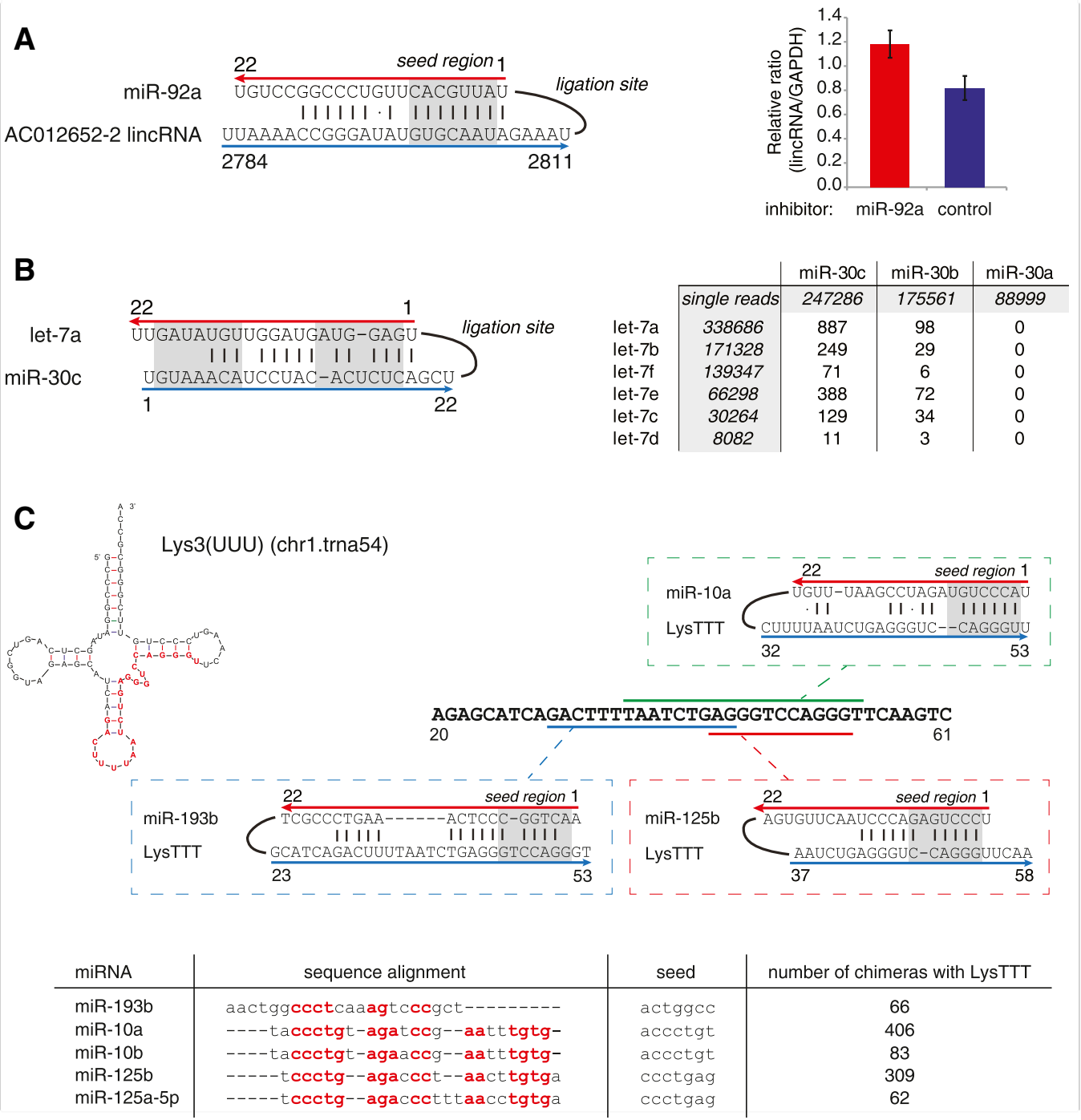

C
Increase in transcript abundance
100.0
80.0
60.0
40.0
20.0
0.0
-20.0
MOTIF
SEED
SCLY
RPL3
RPS5
KAT2A
MAPRE1
C9orf7
C1orf114
MSL3
CCT7
CAPRIN1
TRAP1
CUX1
TANC2
FNIP1
EHBP1
RASAL2
RBM10
PCMTD1
RORA
SRPR
UBXN8
TPRKB
SIP1
ADK
ADAL
D
Cumulative distribution of targets (%)
100
80
60
40
20
0
-0.4
-0.2
0
0.2
0.4
Change in transcript abundance (log2 fold)
transcripts without 7-nt seed
7-nt seed
class IV
class IV + 7-nt seed

Figure 5


transcript enrichment (log 2 fold)
−0.4
−0.2
0.0
0.2
0.4
all genes
NO miR-92 targets
NO miR-92 7-nt seed
miR-92:7-nt seed
all CLASH miR-92 targets
class I
class II
class III
class IV
all
with 7-nt seed
without 7-nt seed
class V
CLASH miR-92a motif

Supp Fig 5




| | miR-30c | miR-30b | miR-30a |
|---|---|---|---|
| | *single reads* | *247286* | *175561* | *88999* |
| let-7a | *338686* | 887 | 98 | 0 |
| let-7b | *171328* | 249 | 29 | 0 |
| let-7f | *139347* | 71 | 6 | 0 |
| let-7e | *66298* | 388 | 72 | 0 |
| let-7c | *30264* | 129 | 34 | 0 |
| let-7d | *8082* | 11 | 3 | 0 |

| miRNA | sequence alignment | seed | number of chimeras with LysTTT |
|---|---|---|---|
| miR-193b | aactgg**ccct**caa**ag**tc**cc**gct--------- | actggcc | 66 |
| miR-10a | ----ta**ccctg**t-**aga**tc**c**g--**aa**ttt**gtg**- | accctgt | 406 |
| miR-10b | ----ta**ccctg**t-**aga**ac**c**g--**aa**ttt**gtg**- | accctgt | 83 |
| miR-125b | -----t**ccctg**--**aga**ccc**t**--**aa**ct**tgtg**a | ccctgag | 309 |
| miR-125a-5p | -----t**ccctg**--**aga**ccctttaa**cc**t**gtg**a | ccctgag | 62 |

**Figure 6**

# cliffs notes

- report **18,500** miRNA-mRNA interactions
- only around **37%** of seed interactions involve **uninterrupted base pairing**
- **16%** of all interactions **do not involve the seed region**
- "competing endogenous RNA" from pseudogenes/lincRNAs **AND abundant stable RNAs** such as tRNA/rRNA to buffer temporary excess of particular miRNAs