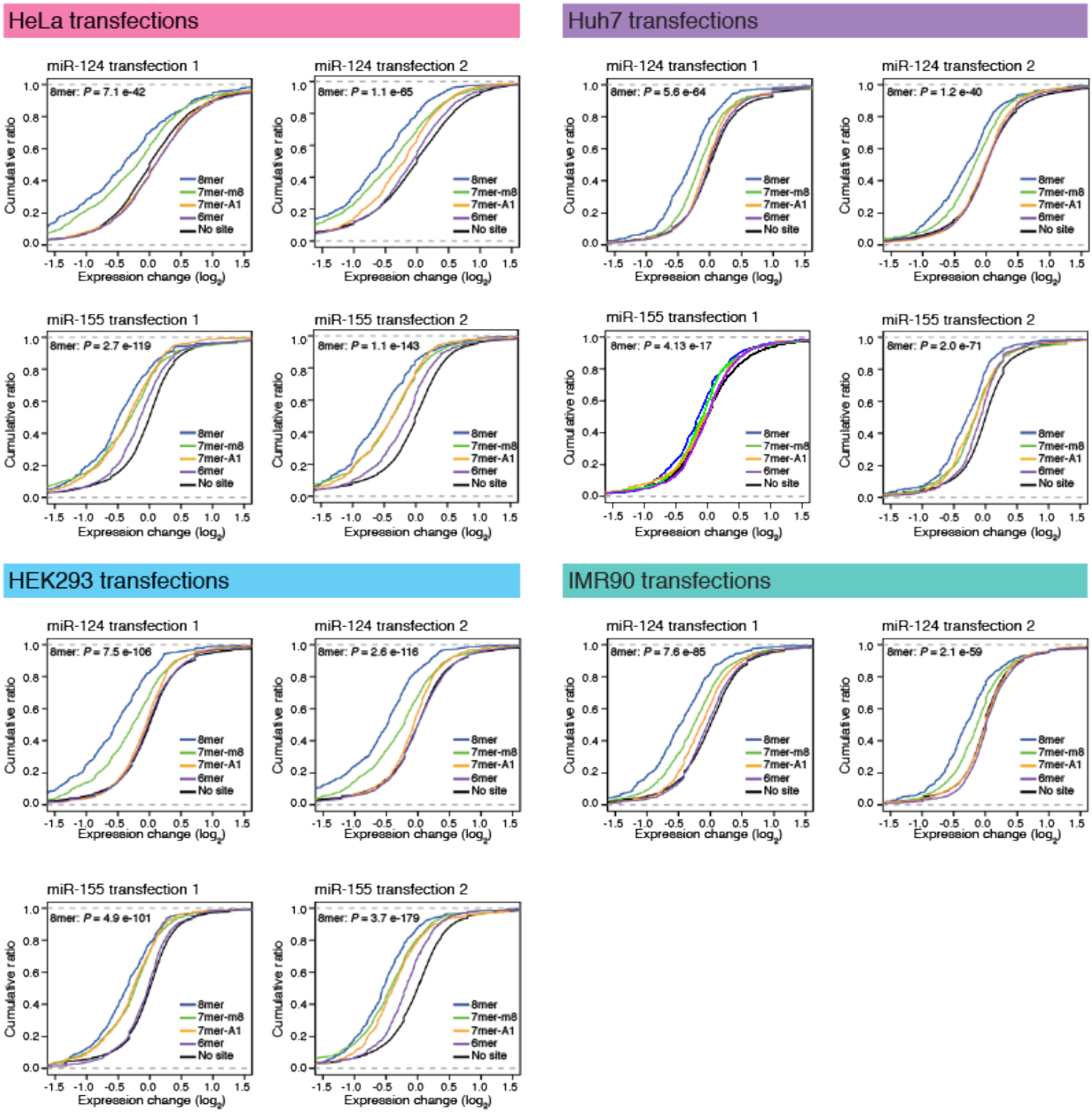
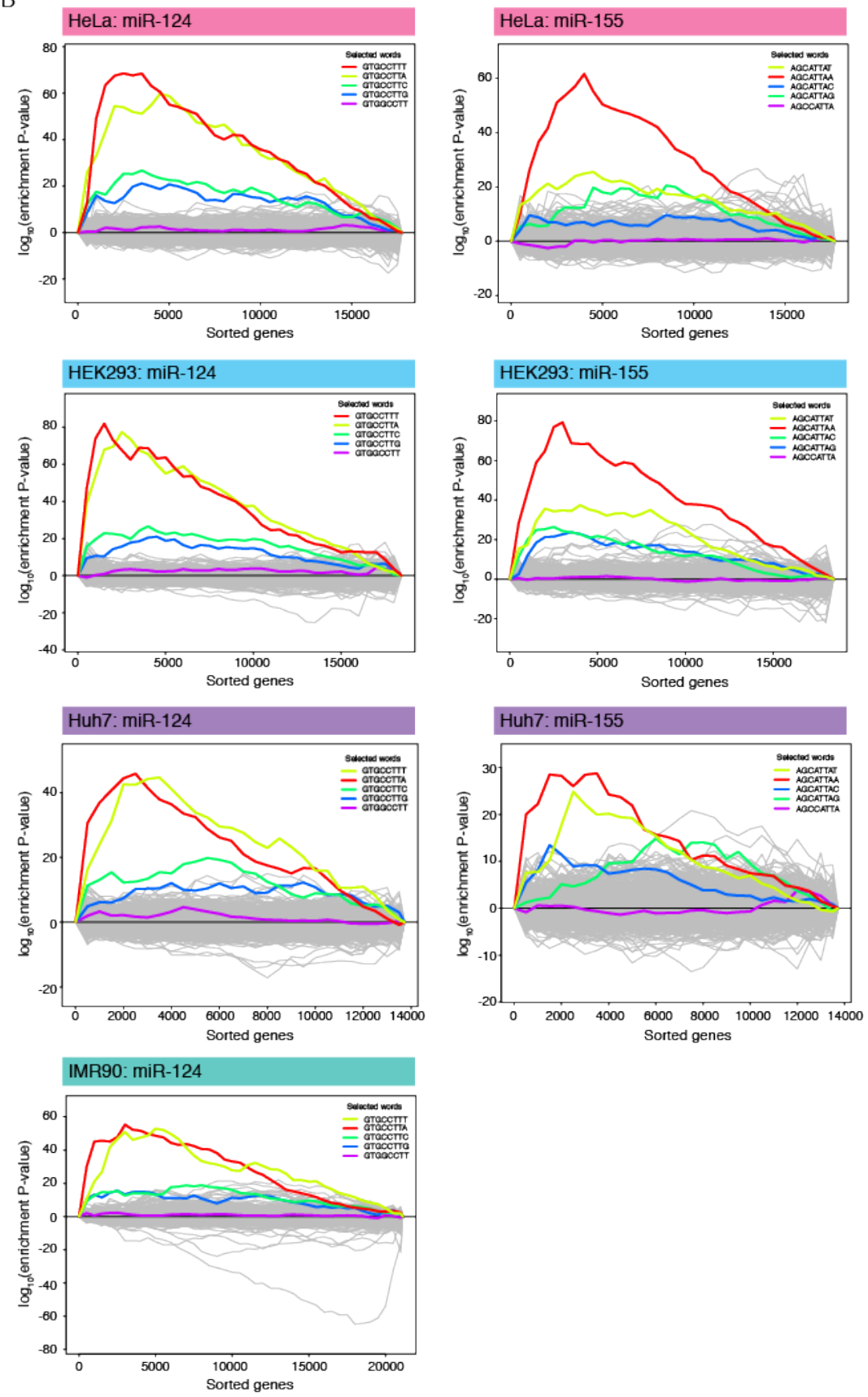
Supplemental Figure 1

A



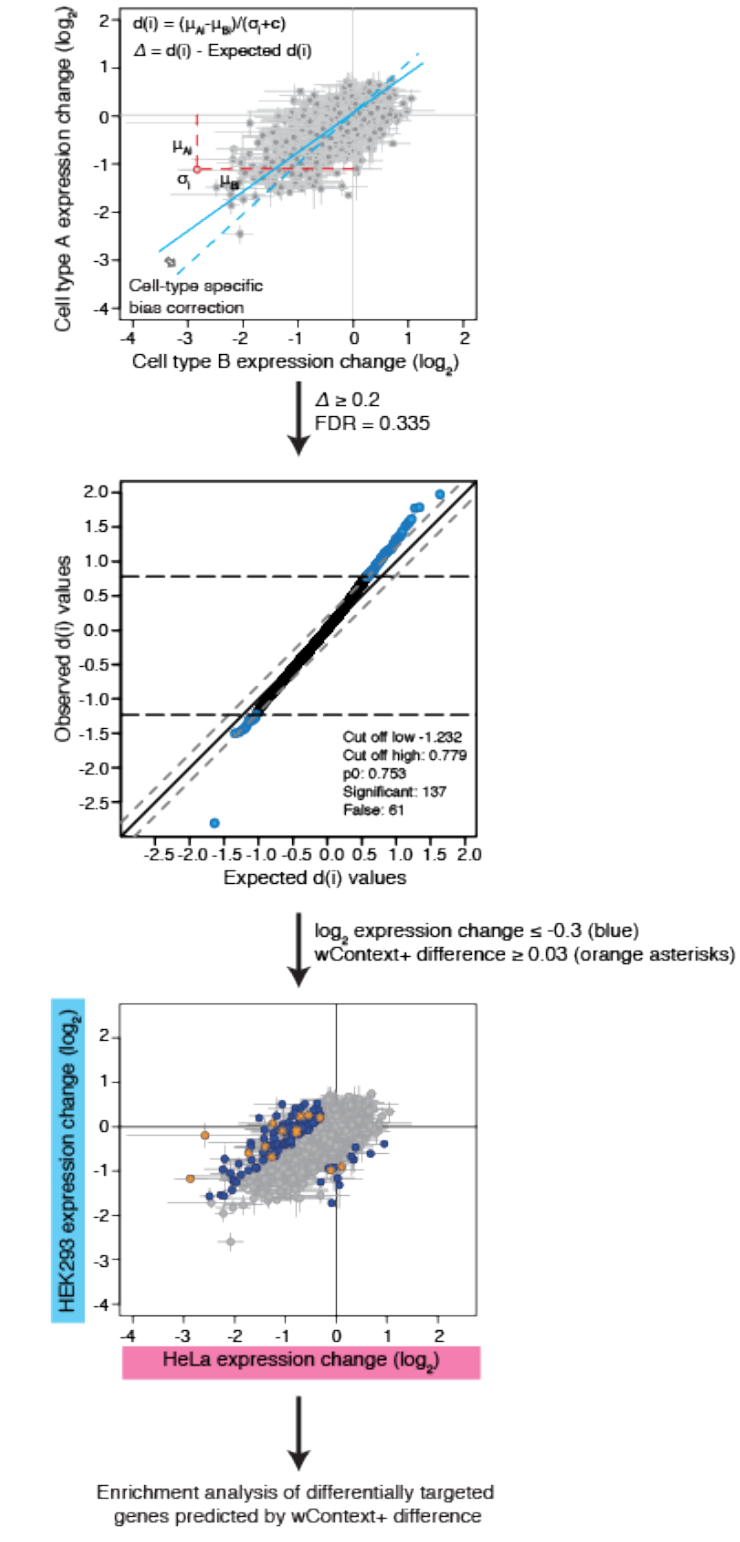
Supplemental Figure 1

B



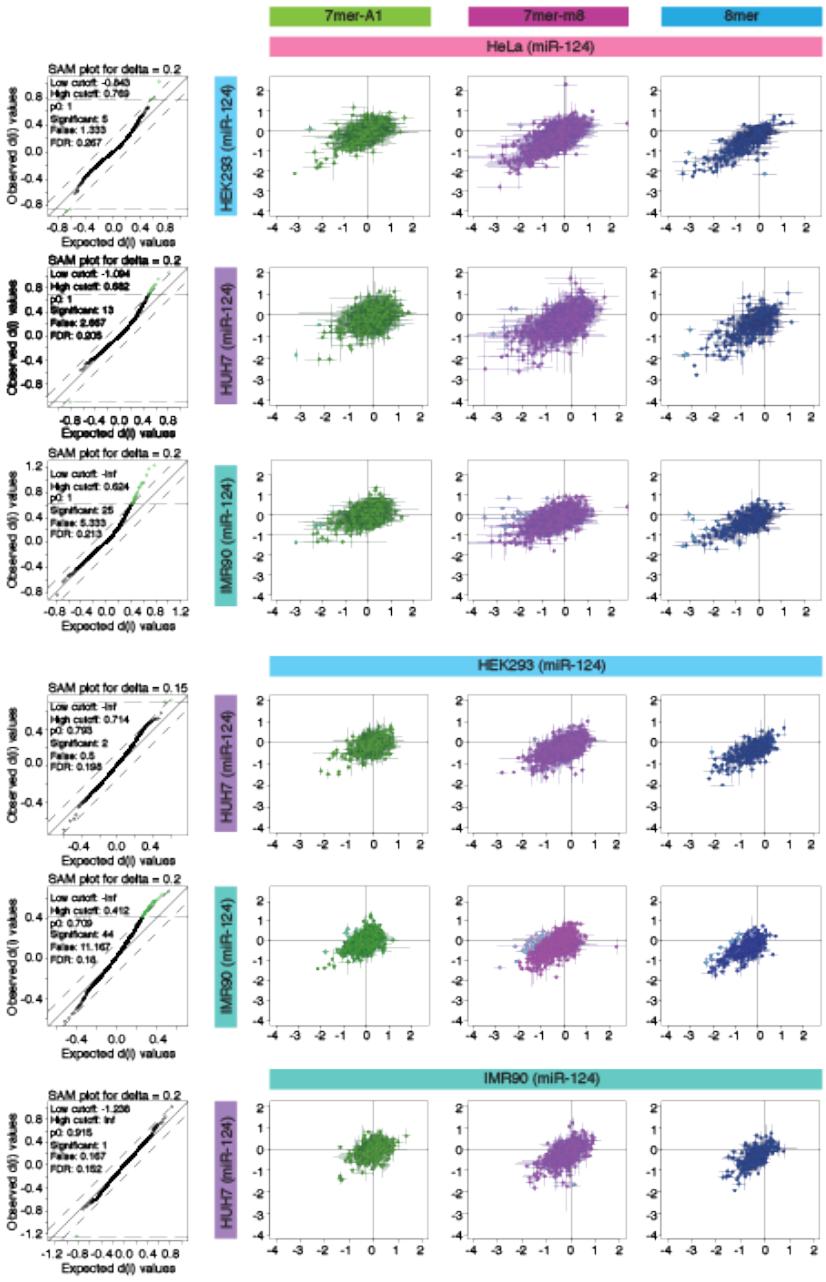
Supplemental Figure 1

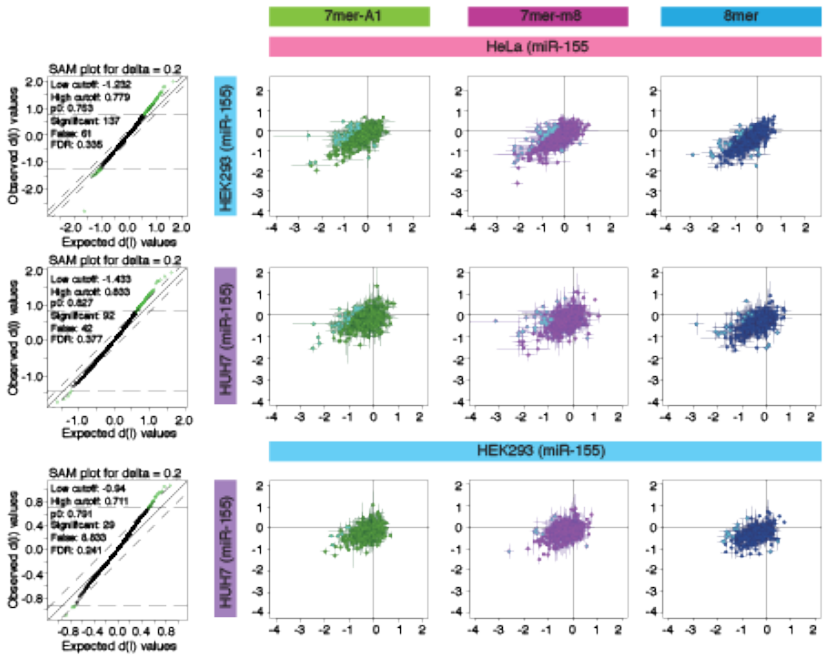
C



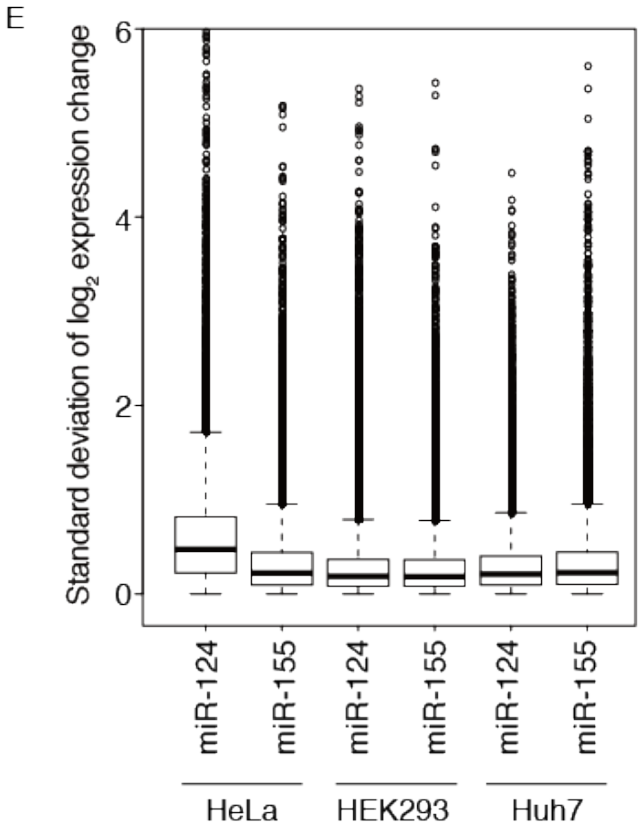
Supplemental Figure 1

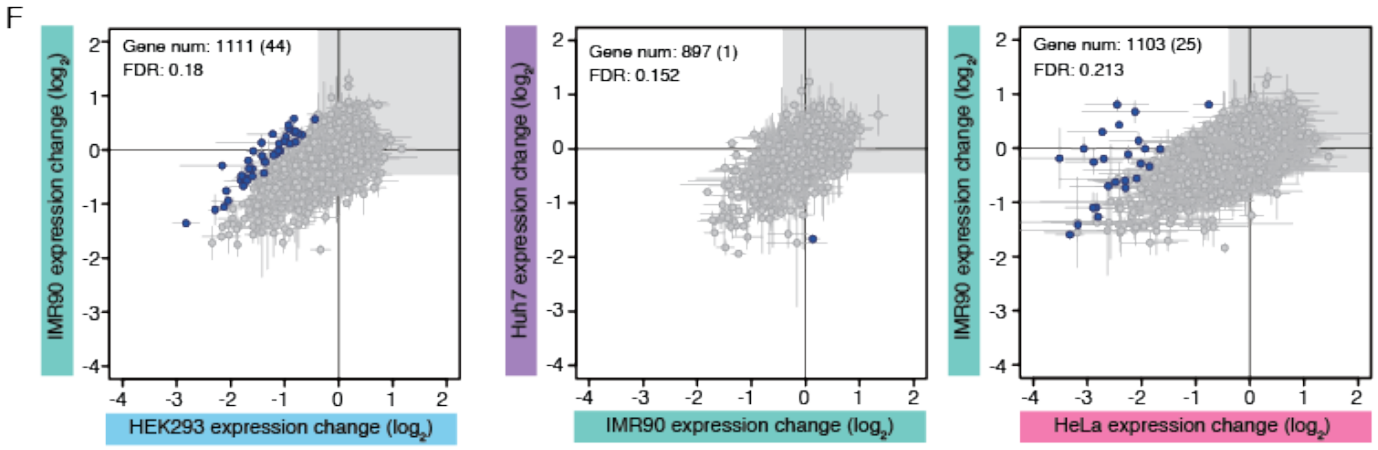
D





Supplemental Figure 1





Supplemental Figure 1, related to Figure 1. miRNA targeting in a variety of cell lines

(A) Cumulative distribution plots showing the fold repression in each miR-124 or miR-155 transfection in HeLa, HEK293, Huh7 and IMR90 cells. For each plot, targets are separated by site type (8mer in blue; 7mer-m8 in green; 7mer-A1 in yellow; 6mer in purple; no site in black). The associated Kolmogorov-Smirnov *P*-value is shown for 8mer sites.

(B) Sylamer analysis (van Dongen et al., 2008) of the transfection combinations shown in (A), with replicates averaged.

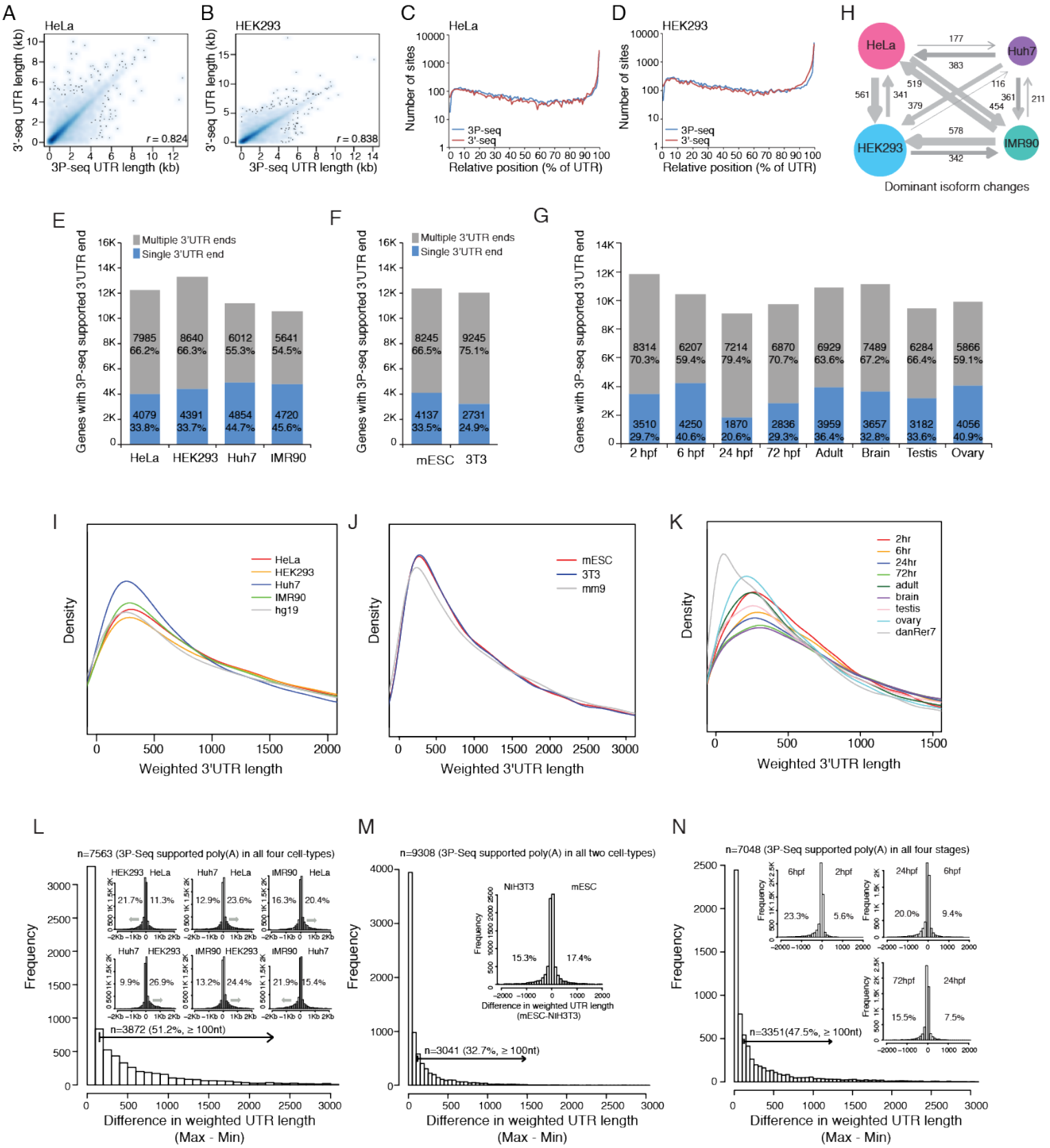
(C) Summary of the strategy to determine differentially repressed targets and those with differential alternative polyadenylation.

(D) All pair-wise comparisons of miRNA-mediated repression for the various miR-124 and miR-155 transfections. For each transfection, targets were subdivided based on site type. Those points in light blue denote genes that were differentially repressed.

(E) Shown are standard deviations of mRNA $\log_2$ changes in miR-124 and miR-155 transfections. mRNAs with 0 standard deviation were excluded in this analysis.

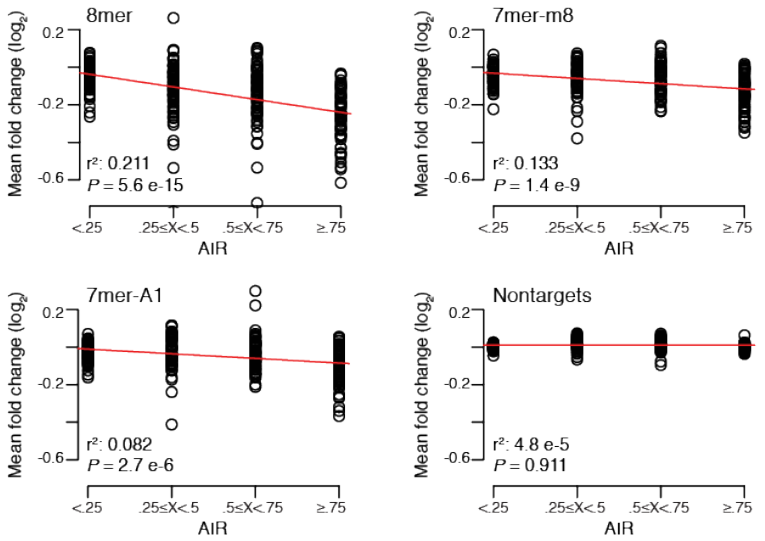
(F) A comparison of the $\log_2$ changes of mRNAs (with at least one 7mer in their 3'UTR) mediated miR-124 in IMR90 cells versus the other three cell types (HeLa, HEK293 and Huh7).

Supplemental Figure 2

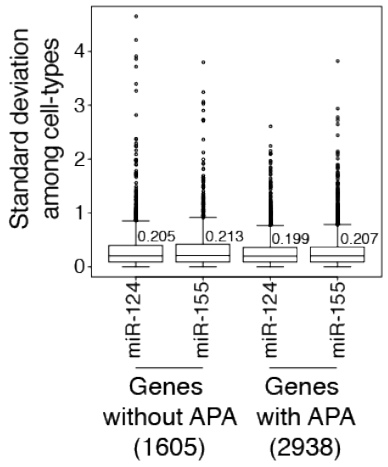


Supplemental Figure 2

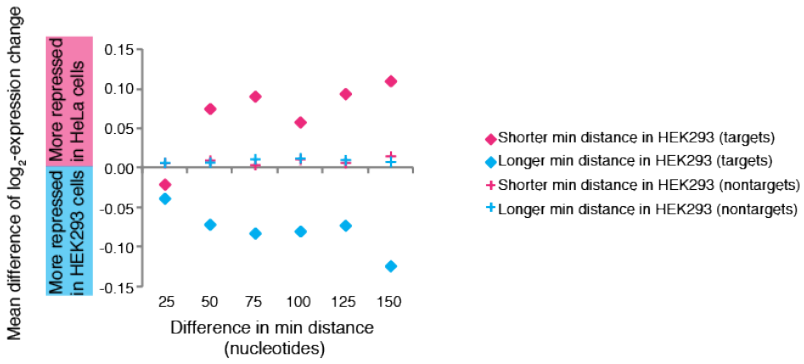
M



N



O



Supplemental Figure 2, related to Figure 2. Poly(A) site usage in a variety of human and mouse cell lines and zebrafish tissues and its effects on miRNA targeting.

(A, B) Poly(A) site usage determined by 3P-seq correlates with that determined with 3'-seq. For genes with more than 10 tags in each method, the weighted 3'UTR length was determined using 3P-seq and 3'-seq (Lianoglou et al., 2013) datasets for HeLa (A) and HEK293 cells (B), the two cell types for which both sets of data were available.

(C, D) The positions of cleavage agree between 3P-seq and 3'-seq. For genes with more than 10 tags in each method, the poly(A) site position (relative to entire length of the longest isoform) is plotted for 3P-seq (blue) and 3'-seq (red) datasets for HeLa (C) and HEK293 (D) cells.

(E–G) The fraction of genes with multiple and single poly(A) sites, as determined by 3P-seq annotation, for four human cell lines (E, HeLa, HEK293, Huh7 and IMR90 cells), two mouse lines (F, mouse embryonic stem cells (mESCs) and NIH3T3 cells), and a variety of zebrafish developmental stages and tissues (G) (Ulitsky et al., 2012).

(H) The number of genes for which the dominant 3'UTR isoform changes between two cell types. The direction of the arrows the represents the cell line with the longer isoform, and the width is proportional to the number of genes changing.

(I–K) Comparison of weighted 3'UTR lengths for (I) the four human cell lines and as annotated in the hg19 reference genome, (J) the two mouse lines and as annotated in the mm9 reference genome, and (K) the zebrafish stages and tissues and as annotated in the danRer7 reference genome.

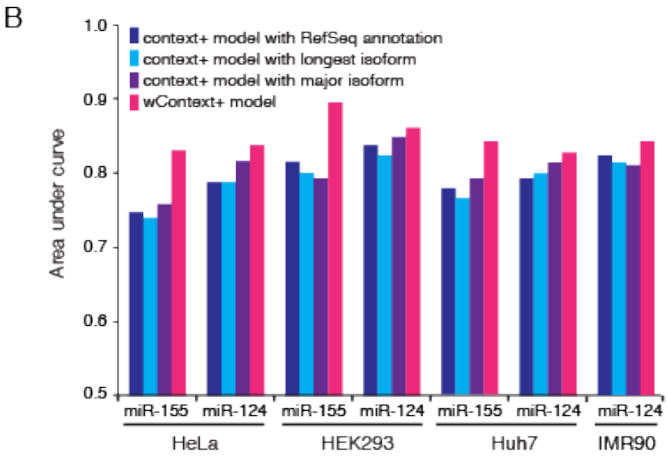
(L-M) Many genes use different poly(A) site in different cellular contexts. Plotted is a histogram of the difference between the longest and shortest weighted 3'UTR lengths for each of the three model organisms. Inset: histograms of various pairwise comparisons.

(O) AIR correlates with miRNA-mediated repression. Using 74 previously published arrays, genes were separated based on their site type and then binned according to their respective AIR. For each bin, the mean fold-change was calculated. The line of best fit is plotted in red.

(P) Variability in expression changes for genes with and without APA. Shown are the standard deviations for the $\log_2$-changes following either miR-124 or miR-155 transfection for genes with and without APA across the four different cell lines (HeLa, HEK293, Huh7 and IMR90). mRNAs with APA are those that have a weighted 3'UTR changing >100 nt between cell types; genes without APA are those whose 3'UTRs change <10 nt. mRNAs with 0 standard deviation were excluded from this analysis. Median values are shown adjacent to the box for each set of genes.

(Q) Changes in minimum distance affect miRNA targeting efficacy. Plotted is the difference in minimum distance between HeLa and HEK293 cells (those with a greater minimum distance in HEK293 cells are in blue; in HeLa cells, red) for targets (diamonds) and non-targets (plus symbols) compared to the difference in repression seen in HeLa and HEK293 cells.

Supplemental Figure 3

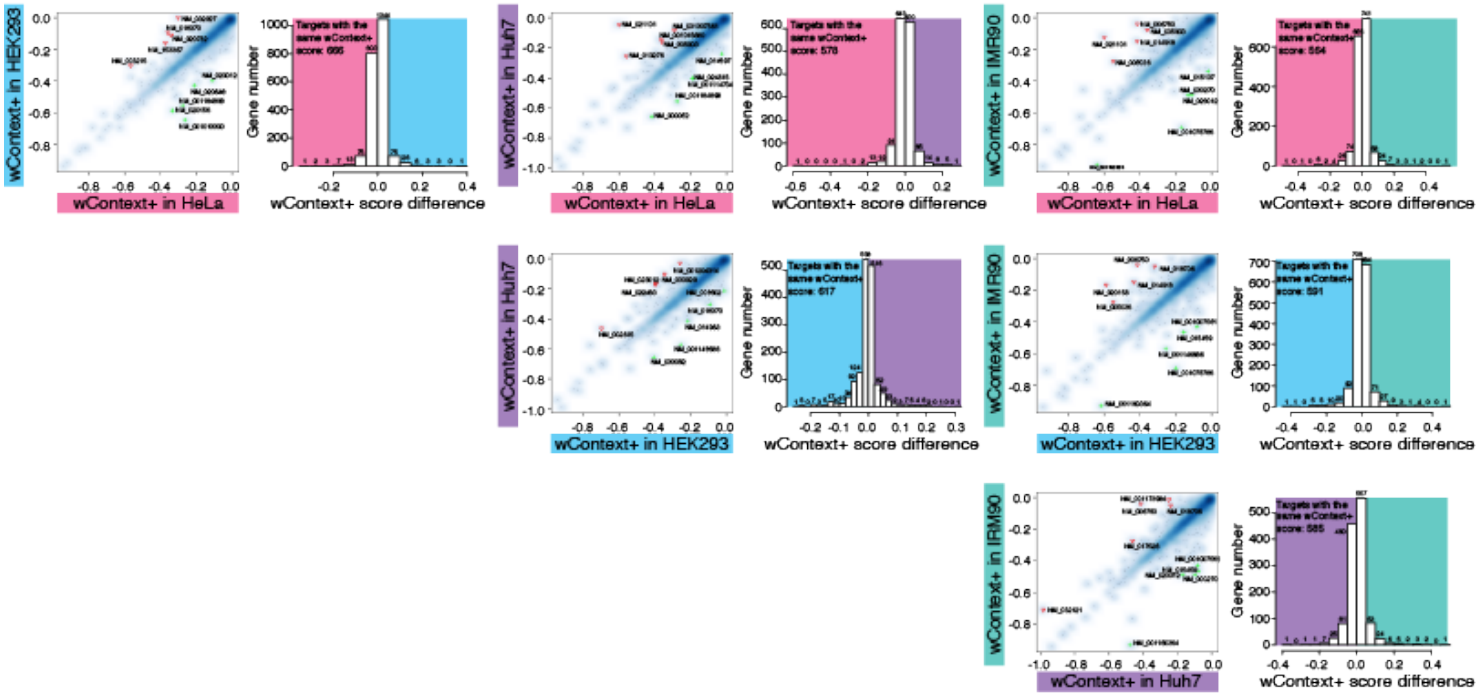


Supplemental Figure 3, related to Figure 3. Evaluation of the performance of the wContext+ model compared to the Context+ model

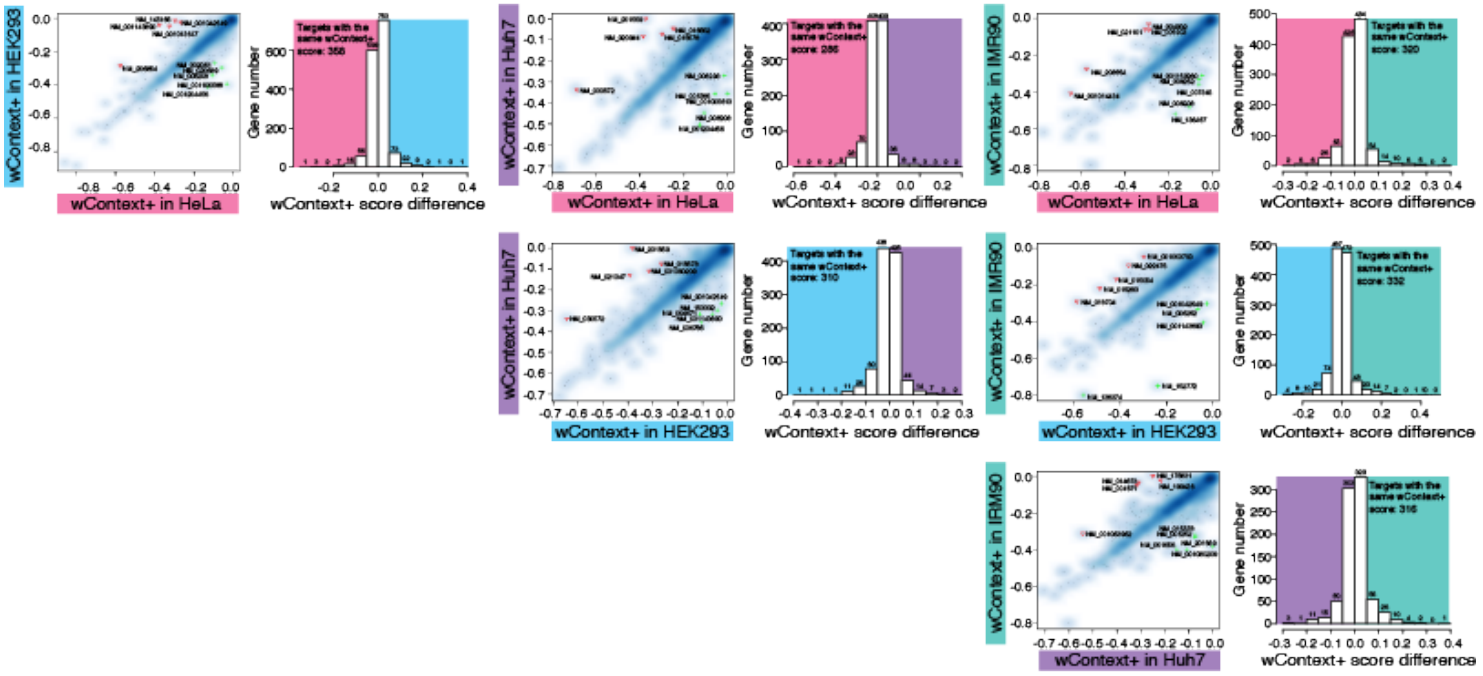
(A) Overview of how alternative polyadenylation affects site inclusion and site efficacy. (B) Performance of various target predictions algorithms was evaluated by the area under the curve. For each transfection of miR-124 and miR-155 into the various human cell lines (HeLa, HEK293, Huh7 and IMR90), the wContext+ model was compared to the Context+ model with RefSeq annotations (blue), the longest 3'UTR annotated with 3P-seq (red) and the major 3'UTR annotated with 3P-seq (green) .

Supplemental Figure 4

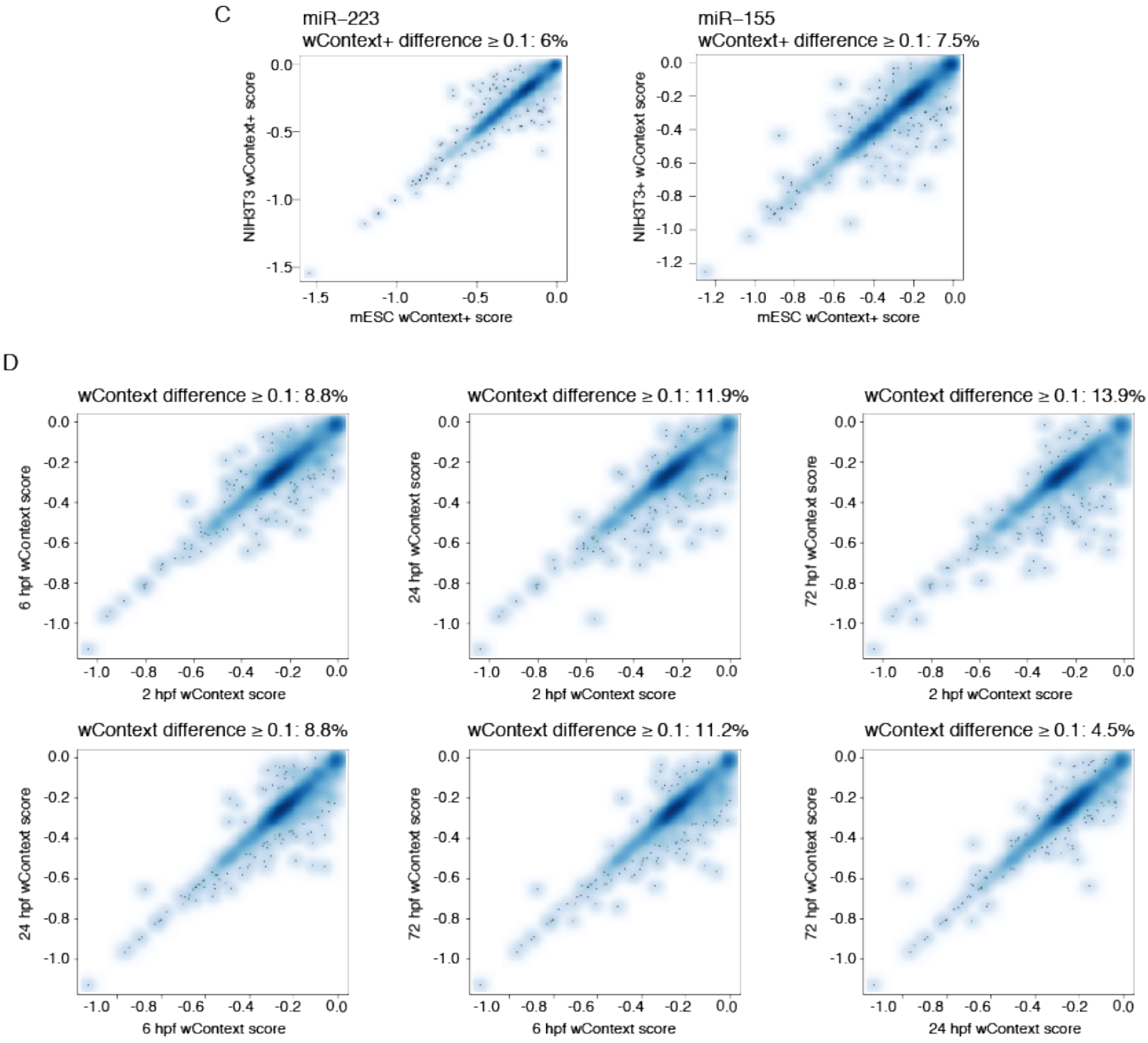
A miR-124 targets

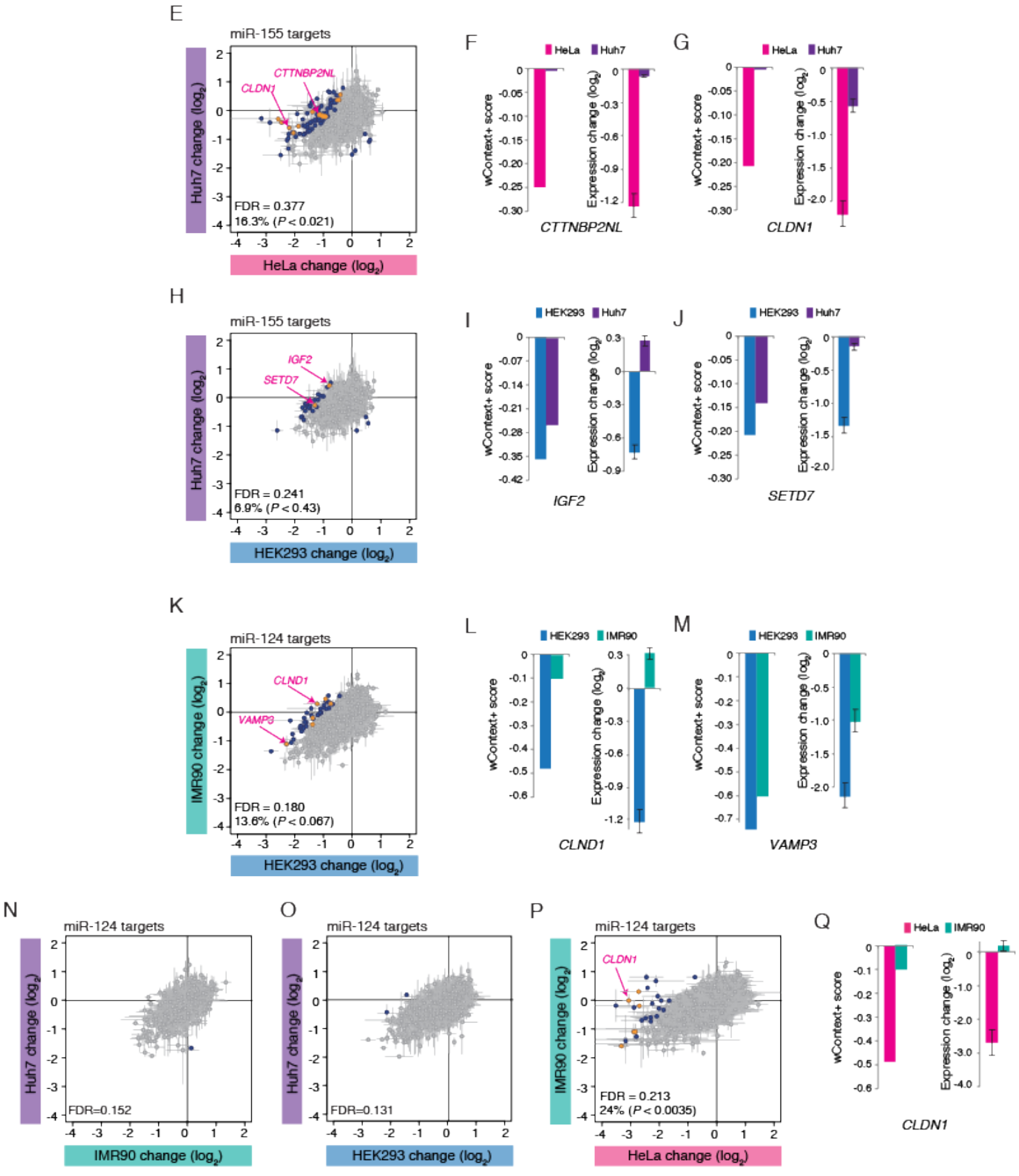


B miR-155 targets



Supplemental Figure 4




E
miR-155 targets
CTTNBP2NL
CLDN1
Huh7 change (log2)
HeLa change (log2)
FDR = 0.377
16.3% (P < 0.021)
F
HeLa
Huh7
wContext+ score
Expression change (log2)
CTTNBP2NL
G
CLDN1
H
miR-155 targets
IGF2
SETD7
FDR = 0.241
6.9% (P < 0.43)
HEK293 change (log2)
I
HEK293
IGF2
J
SETD7
K
miR-124 targets
CLND1
VAMP3
IMR90 change (log2)
FDR = 0.180
13.6% (P < 0.067)
L
IMR90
CLND1
M
VAMP3
N
miR-124 targets
FDR=0.152
O
FDR=0.131
P
FDR = 0.213
24% (P < 0.0035)
Q
CLDN1

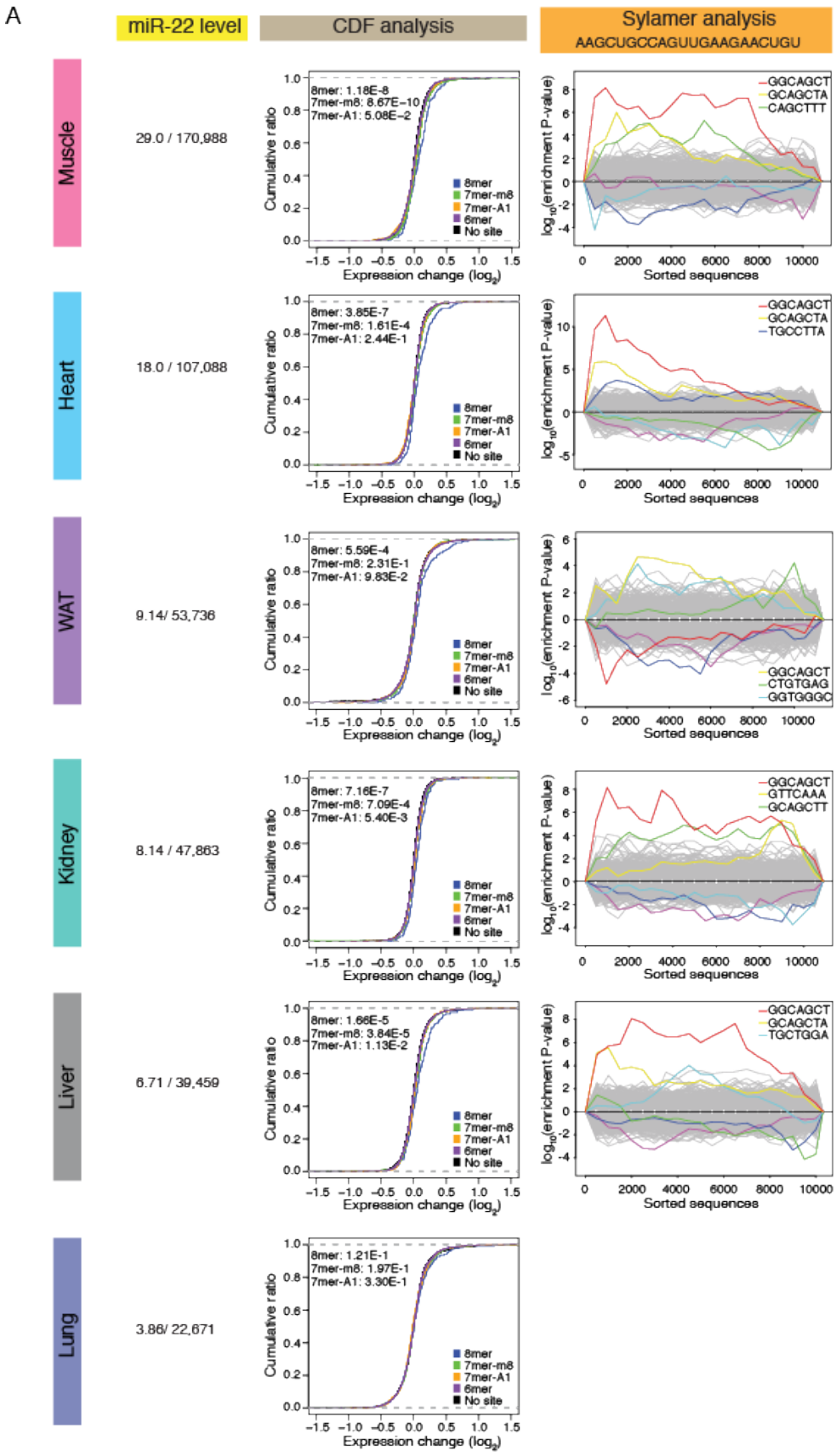
Supplemental Figure 4, related to Figure 4. The effect of cell type upon the wContext+ score

For each pair of the four human cell lines (HeLa, HEK293, Huh7 and IMR90), the wContext+ scores of predicted miR-124 (A) and miR-155 (B) targets are compared.

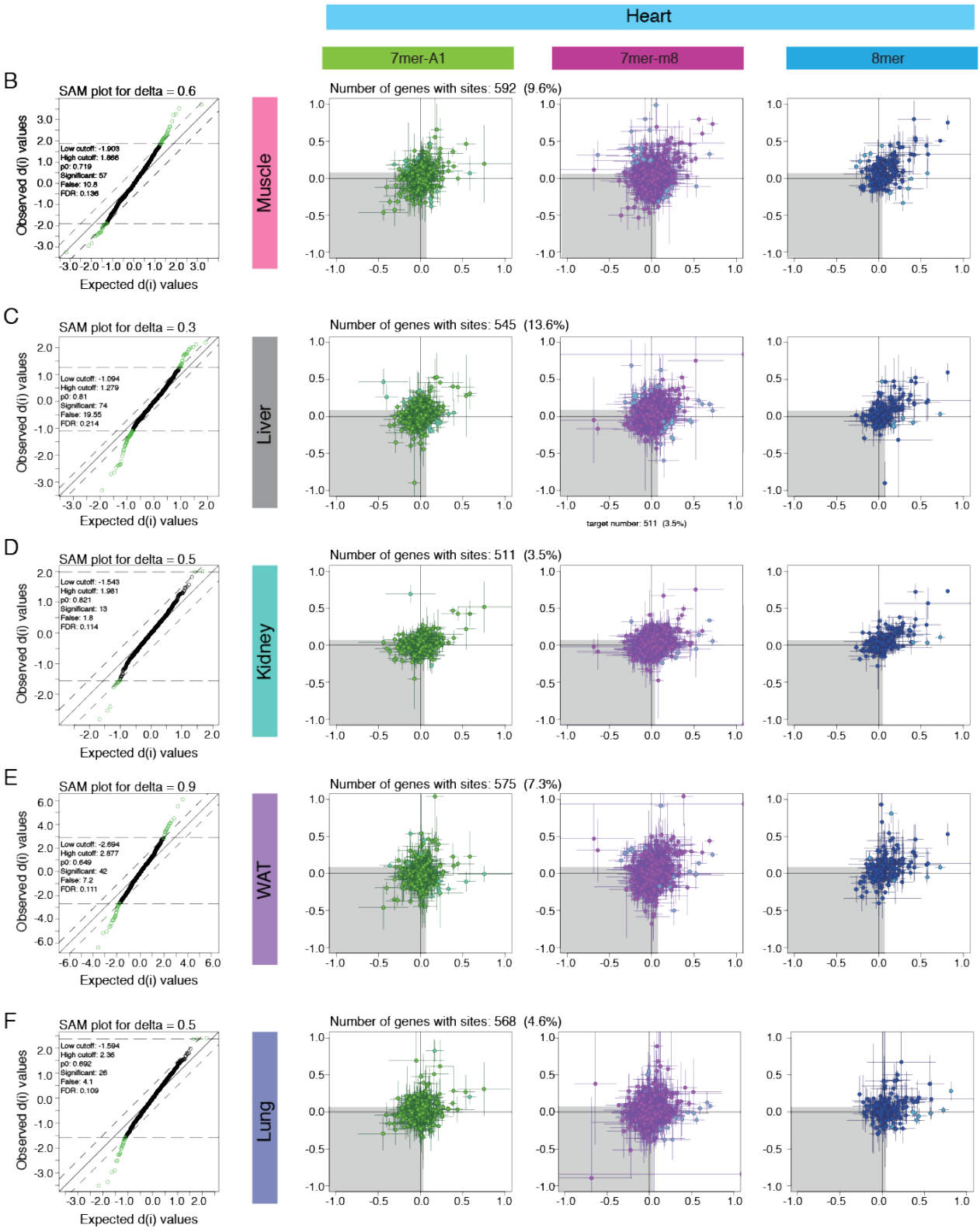
(C-D) For each pair of mouse cell types (C) or zebrafish developmental stages (D), the wContext+ scores of predicted miRNA targets (miR-223 and miR-155 for mouse (C) and miR-430 for zebrafish (D)) are compared.

(E-Q) As in Figure 4, genes with differential 3'UTR usage are enriched in genes that are differentially repressed. Dots in blue are those whose repression is significantly different between the two lines (with $\Delta \geq 0.2$). Those in yellow are genes whose repression is significantly different and whose wContext+ score differs between the two lines (wContext+ score difference ≥ 0.02 (E), or ≥ 0.03 (H, K, N, O, P)).

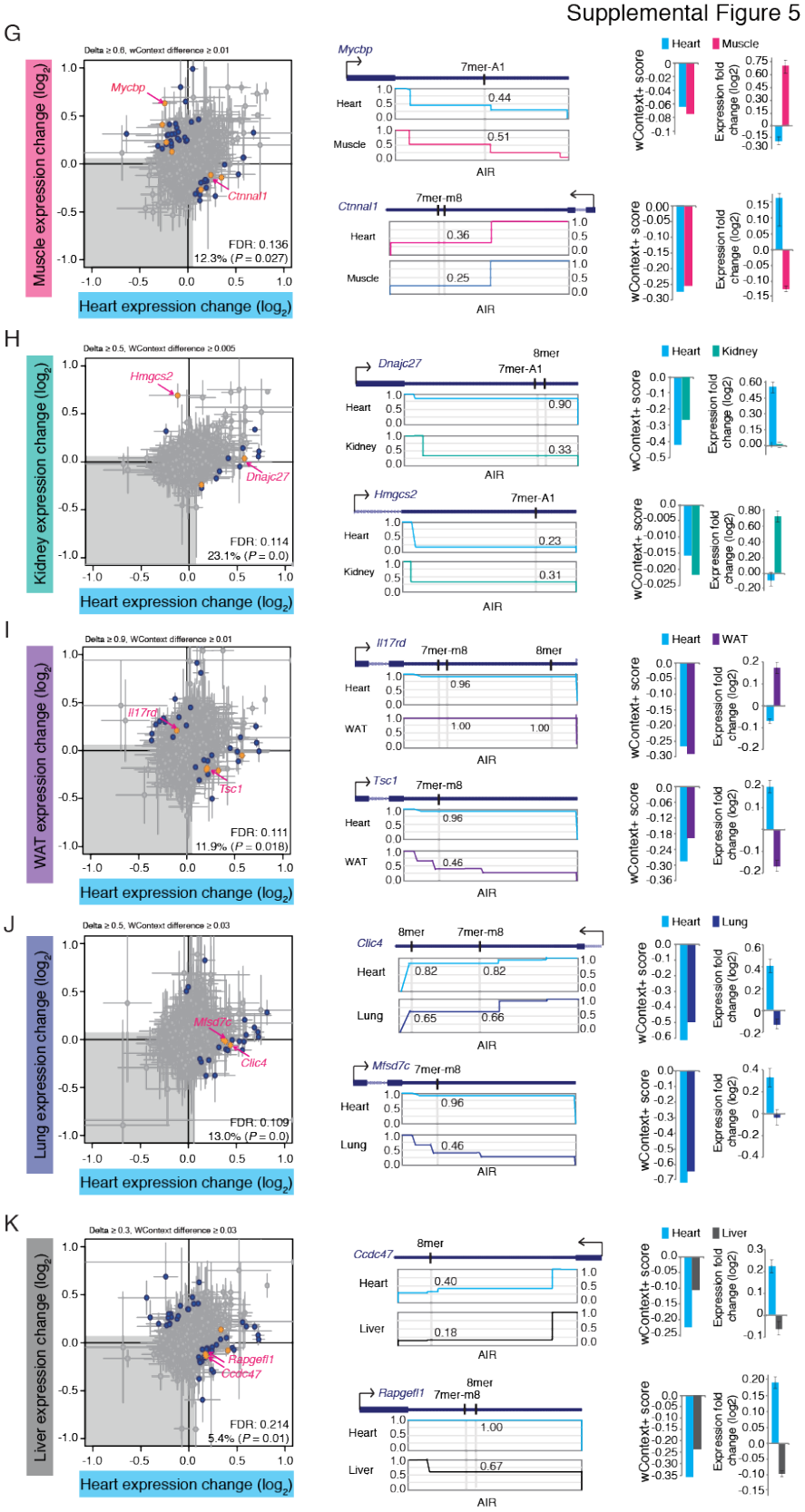
Supplemental Figure 5



Supplemental Figure 5



Supplemental Figure 5



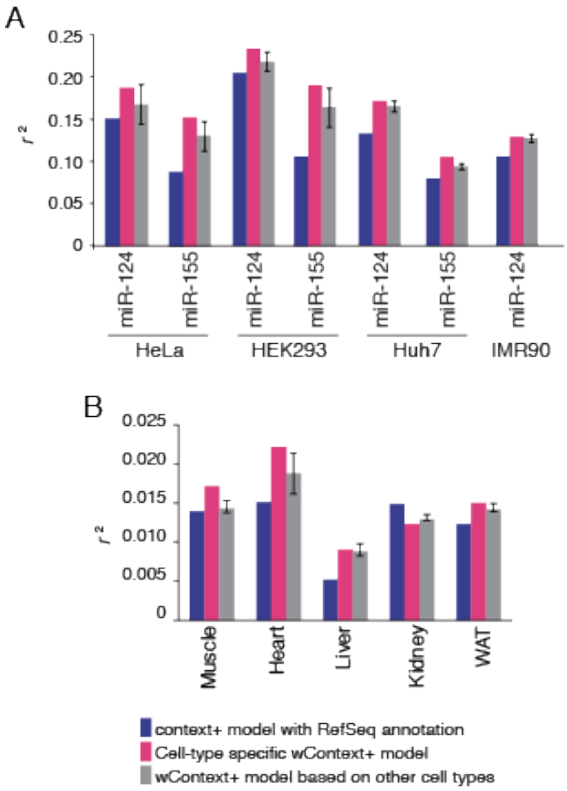
Supplemental Figure 5, related to Figure 5. The differential effect mediated by miRNAs in their endogenous contexts

(A) For each tissue, the repression mediated by miR-22 was calculated by comparing RNA from wild-type mice and those lacking miR-22. The relative level and estimated reads were calculated based on data from testes. The cumulative distribution for targets (separated by site type) is plotted (left panel). Significance of targeting was determined using the Kolmogorov-Smirnov test. Sylamer analysis was also performed for each tissue (right panel).

(D–H) As in Figure S1D, repression mediated by miR-22 in heart is compared to that in the other five tissues profiled (muscle, liver, kidney, WAT and lung). For each comparison, targets were subdivided based on site type. Those points in light blue denote genes that were differentially repressed.

(I–M) As in Figure 1, the repression mediated by miR-22 in heart and other tissues, as determined by comparing wild-type mice and those lacking miR-22, is plotted for mRNAs containing sites in their 3'UTRs. Those dots in blue represent targets displaying differential repression ($\Delta \geq 0.6$). Those in yellow are targets that are differentially repressed and have different wContext+ scores (wContext+ score difference ≥ 0.01). The region corresponding to a $\log_2$ change of less than 0.1 is shaded in grey. Representative examples of targets with differential 3'UTR isoform usage are given for each comparison.

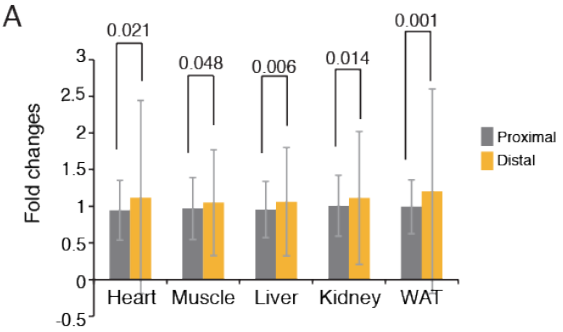
Supplemental Figure 6



Supplemental Figure 6, related to Figure 6. Cell-type–specific wContext+ models outperform non-cell-type–specific models, even when differential targets are excluded

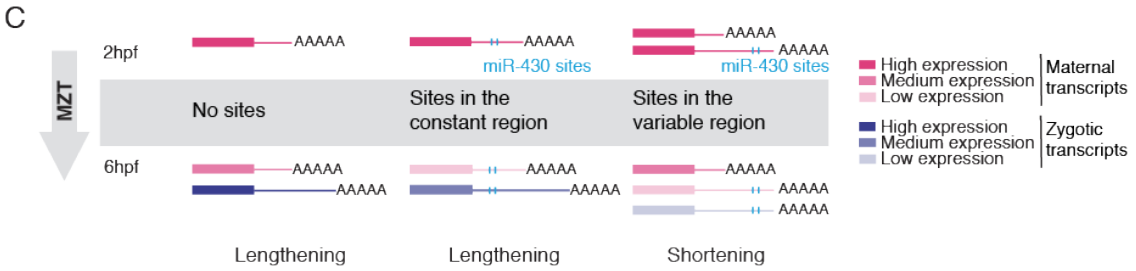
As in Figure 6, the performance of non-cell-type–specific wContext+ models for exogenous (A) and endogenous miRNAs (B) is compared with that of cell-type–specific wContext+ models; however, here, those differential targets whose differential repression can be explained by differential alternative polyadenylation are excluded. A comparison of performance of the original context+ model (dark blue), the cell-type–specific wContext+ model (pink) and the wContext+ model based on 3P-seq from other cell types (grey).

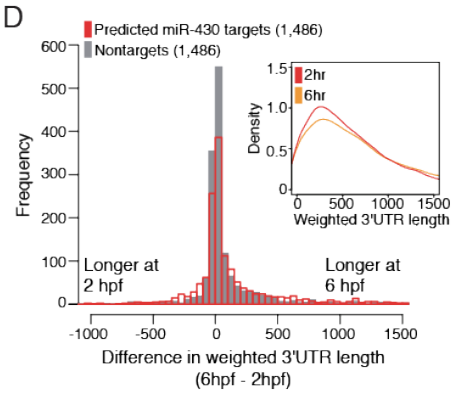
Supplemental Figure 7

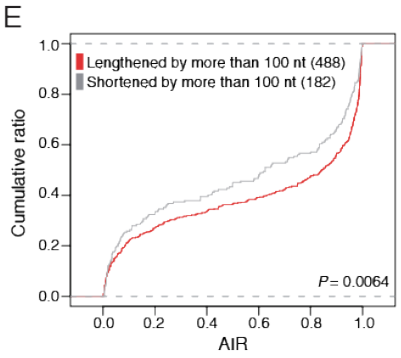


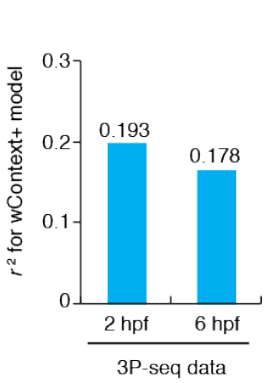
B

	Heart	Muscle	Liver	Kidney	WAT
P values	0.021	0.048	0.006	0.014	0.001
Random cohorts	0/100	5/100	0/100	8/100	0/100









Supplemental Figure 7, related to Figure 7. miRNA targeting shapes the 3'UTR landscape.

(A) miR-22 targeting changes 3'UTR isoform usage. To determine the effect of miR-22 on the expression of 3'UTR isoforms, genes with sites in the 3'UTR variable region (defined as an AIR between 0.1 and 0.9) were examined (3P-seq tags > 10). For five tissues (heart, muscle, liver, kidney and white adipose tissue (WAT)), the fold changes of isoforms using proximal poly(A) sites (grey) and distal poly(A) sites (orange) were calculated by comparing 3P-seq data from wild-type and *miR-22Δ* mice. Plotted are the mean fold-changes and their standard deviations (A).

(B) As in (A), but for 100 control gene cohorts containing a control site in the variable region. The number of times where the changes of proximal and distal isoforms varied more significantly than that for the miR-22 targets is indicated.

(C) During the maternal-zygotic transition (MZT), the 3'UTR landscape is determined by zygotic transcription (which generally lengthens UTR length) and targeting by the major embryonic miRNA, miR-430, which is responsible for clearing many maternally inherited messages (Giraldez et al., 2006). The combined effects of these processes differ depending on the position of the miR-430 site.

(D) The 3'UTRs of miR-430 targets are affected during MZT. The difference in weighted 3'UTR length for nontargets (in grey) and miR-430 (red), comparing between 2 and 6 hours post fertilization (hpf). Shown in the inset is the weighted 3'UTR length for genes at 2 and 6 hpf.

(E) The targets whose 3'UTRs shorten during MZT tend to have a higher AIR at 2 hpf. For the 783 genes with miR-430 sites in their constant regions (AIR ≥ 0.95), we observed

increased accumulation of longer UTR isoforms, presumably because zygotic transcription resulted in overall 3'UTR lengthening. In contrast, for the 1,538 genes with sites in the variable region, we observed more of the shorter isoform, presumably because the long maternal and zygotic isoforms are targeted by miR-430, and thus at 6 hpf these mRNAs have shorter 3'UTR. Plotted are the AIRs for predicted targets whose 3'UTRs lengthened or shortened by more than 100 nucleotides during MZT. Those that lengthened had more miR-430 sites in the 3'UTR constant region (i.e. had a higher AIR) those whose 3'UTRs shortened (~38% vs. ~24 % sites had AIR ≥ 0.95, $P = 0.0064$).
(F) Using poly(A) site annotation at 2 hpf more accurately predicts miR-430 repression in zebrafish embryos. A comparison of the performance of the wContext+ model based on 3P-seq data from either 2 hpf or 6 hpf.

SUPPLEMENTAL TABLES

Supplemental Table 1. Number of significantly differential targets in the human cell line transfections.

		miR-124				miR-155			
		Targets*	**SDT**	**SDT_{FDR}**	**Percent**	**Targets***	**SDT**	**SDT_{FDR}**	**Percent**
HeLa	**HEK293**	1169	4	2.9	0.3%	991	137	91.1	9.2%
	Huh7	1098	13	10.4	0.9%	921	92	57.3	6.2%
	IMR90	1103	25	19.7	1.8%	N/A	N/A	N/A	N/A
HEK293	**Huh7**	1164	2	1.7	0.1%	1013	29	22.0	2.2%
	IMR90	1111	44	36.1	3.2%	N/A	N/A	N/A	N/A
Huh7	**IMR90**	897	1	0.9	0.1%	N/A	N/A	N/A	N/A
Average		1090.3	**14.8**	**11.9**	**1.1%**	**975.0**	**86.0**	**56.8**	**5.9%**

*: $\log_2$ fold change < -0.3 in either sample
SDT: significantly differential targets
SDT_{FDR}: SDT adjusted by false discovery rate (FDR)
Percent: SDT_{FDR}/Targets

Supplemental Table 2. Mapping statistics for 3P-seq libraries.

A. Uniquely mapped 3p-seq tags for the human cell lines.

Sample	Uniquely mapped tags
HeLa	5536210
HEK293	4375957
Huh7	4707886
IMR90	3340695
Total	17960748

B. Statistics of 3P-seq clusters and tags.

Sample	Clusters in 3'UTR	Clusters not in 3'UTR	Fraction	Cluster tags in 3'UTR	Cluster tags not in 3'UTR	Fraction
HeLa	35393	14430	0.71	3623308	525821	0.87
HEK293	42861	19565	0.69	2923958	380549	0.88
Huh7	30587	13708	0.69	1786085	1035002	0.63
IMR90	22421	6504	0.78	2470059	151337	0.94

C. Comparison of the major 3'UTR isoform (based on 3P-seq data) with RefSeq annotation in human cell lines.

Sample	Longer	Shorter	No Change	Longer	Shorter	No Change
HeLa	1073	2980	5721	11.0%	30.5%	58.5%
HEK293	890	3033	5680	9.3%	31.6%	59.1%
Huh7	743	2679	4785	9.1%	32.6%	58.3%
IMR90	477	2096	4366	6.9%	30.2%	62.9%

D. Uniquely mapped 3P-seq tags for mESCs and NIH3T3 cells.

Sample	Uniquely mapped tags
mESC	4743687
NIH3T3	5751050
Total	10494737

E. Statistics of 3P-seq clusters and tags for mESCs and NIH3T3 cells.

Sample	Clusters in 3'UTR	Clusters not in 3'UTR	Fraction	Cluster tags in 3'UTR	Cluster tags not in 3'UTR	Fraction
mESC	40993	44447	0.48	3207024	465646	0.87
3T3	21156	25445	0.45	3859342	734798	0.84

F. Comparison of the major 3'UTR isoform (based on 3P-seq data) with RefSeq annotation in mouse cell lines.

Sample	Longer	Shorter	No Change	Longer	Shorter	No Change
mESC	1754	3378	7224	14.2%	27.3%	58.5%
3T3	1452	3041	6483	13.2%	27.7%	59.1%

G. Uniquely mapped 3P-seq tags for the zebrafish samples.

Sample	Uniquely mapped tags
2hpf	4743687
6hpf	5751050
24hpf	7423123
72hpf	9813474
Adult	7470432
Brain	7905265
Ovary	9233031
Testis	7621156
Total	59961218

H. Statistics of 3P-seq clusters and tags for the zebrafish samples.

Sample	Clusters in 3'UTR	Clusters not in 3'UTR	Fraction	Cluster tags in 3'UTR	Cluster tags not in 3'UTR	Fraction
2hpf	63719	54762	0.54	3661416	1832790	0.67
6hpf	40595	53755	0.43	3499240	1551141	0.69
24hpf	23268	28626	0.45	4108748	1699501	0.71
72hpf	27090	41024	0.40	5155659	2379962	0.68
Adult	33267	38779	0.46	3804244	1907906	0.67
Brain	25057	42613	0.37	3144426	3705035	0.46
Ovary	32024	36822	0.47	4919208	2257576	0.69
Testis	35304	92435	0.28	3673964	2392767	0.61

I. Comparison of the major 3'UTR isoform (based on 3P-seq data) with RefSeq annotation in zebrafish.

Sample	Longer	Shorter	No Change	Longer	Shorter	No Change
2hpf	1605	966	4427	22.9%	13.8%	63.3%
6hpf	1894	693	4392	27.1%	9.9%	62.9%
24hpf	2446	791	4620	31.1%	10.1%	58.8%
72hpf	3107	766	5097	34.6%	8.5%	56.8%
Adult	2138	1279	5011	25.4%	15.2%	59.5%
Brain	2724	523	4780	33.9%	6.5%	59.5%
Ovary	1508	1260	4359	21.2%	17.7%	61.2%
Testis	2860	1125	5495	30.2%	11.9%	58.0%

Supplemental Table 3. Genes with differential wContext+ scores are differentially targeted in human cells.

For each pairwise comparison of cell types, the fold-changes for genes containing miR-124 or miR-155 sites that have different wContext+ scores are listed.

Supplemental Table 4. Mapping statistics for 3P-seq libraries from mouse tissues.

Sample	Uniquely mapped tags
Heart (KO)	4713978
Heart (WT)	4549578
Muscle (KO)	3181063
Muscle (WT)	5534321
Liver (KO)	4314908
Liver (WT)	5519789
Kidney (KO)	3812740
Kidney (WT)	4892026
Lung (KO)	8669132
Lung (WT)	6916098
WAT (KO)	2710405
WAT (WT)	3734765

Supplemental Table 5. Genes with differential wContext+ scores are differentially targeted by endogenous miR-22.

For each pairwise comparison of tissues, the fold-changes for genes containing miR-22 sites that have different wContext+ scores are listed.

SUPPLEMENTAL MATERIALS AND METHODS

Data sources. Human genome assembly hg19, mouse genome assembly mm9, and zebrafish genome assembly danRer7 were used throughout the study. For annotated mRNA sets, NCBI RefSeq gene annotations (hg19:Aug-22-2011, mm9:Mar-29-2009, and danRer7: Oct-3-2011) were used. Compiled microarray data for 74 mi/siRNA transfections were from Garcia et al. (2011). Microarray data were from the NCGI GEO: for miR-223 knockout and wild-type cells, GSE22003 (Guo et al., 2010); for miR-155 knockout and wild-type cells, E-TABM-232; for zebrafish wild-type and MZDicer mutant, GSE4201 (Giraldez et al., 2006). The microarray data for miR-22 knockout and wild-type samples are available from NCGI GEO (GSE52940). The RNA-seq and 3P-seq datasets are available from NCGI GEO (GSE52531).

Plasmid construction. To generate luciferase reporters, ~500-800 nucleotides of 3'UTRs containing the miR-155 site from predicted differential targets were amplified from genomic DNA using nested PCR. These were digested as indicated and ligated into digested pIS1. The miR-155 site(s) were then mutated with the indicated oligonucleotides using the QuikChange mutagenesis kit or QuikChangeMulti mutagenesis kit, as indicated by the manufacturer (Agilent).

HK1. Primers for the first PCR: GCGAGAACAGAGGACTGGAC and GATGCACGCCTGTAGTCTCA. Primers for the second PCR: CTAGTCTAGA GAGTCCGGGATCCCCAGCCTACTG (containing XbaI site) and GCGCACTAGCGGCCGC CGACACGGCTCACAAAGCGGTGGG (containing NotI site). Mutation oligonucleotides: CTGAAGGCGAGTGTGGGCATAaCgTaAGCTGCTTCCTCCCCTCCTG and CAGGAGGGGAGGAAGCAGCTtAcGtTATGCCCACACTCGCCTTCAG.

ATP2B2. Primers for the first PCR: TGTTAGGGCCCAGATTTCAC and GAGGACAGTGGGTTGGAGAG. Primers for the second PCR: CTAGTCTAGA GGGTGGTGGACTATCTTGAAGAACCC (containing XbaI site) and GCGCACTAGCGGCCGC CTGACATCTGTTATGAAAATTACC (containing NotI site).

Mutation oligonucleotides:

CTGTAATTTCTACCAGAAATTTCCAGAaCgTaATGTAGGTAGAAAAAAATGCAAGCAA

GC and

GCTTGCTTGCATTTTTTTCTACCTACATtAcGtTCTGGAAATTTCTGGTAGAAATTACAG

.

PELO. Primers for the first PCR: AGCTGCAAGAATGGCACTTT and

GGAATGACTTCCAGCTTTGC. Primers for the second PCR: CTAGTCTAGA

CACACTGCATGCTTTGAAACAG (containing XbaI site) and GCGCACTAGCGGCCGC

GCCTCAAGGAGATTATTTTGAC (containing NotI site). Mutation oligonucleotides:

GGAAATTCTTGGTTTTGGTTTGTTGTTGTTAaCgTaAGTCATATTTGATTTAGAGGGTA

ACTTAAATC and

GATTTAAGTTACCCTCTAAATCAAATATGACTtAcGtTAACAACAACAAACCAAAACC

AAGAATTTCC.

INSIG2. Primers for the first PCR: ATTGCCAGGATGCAGTTTTC and

ATCTTGCCCTTGTGGATCAG. Primers for the second PCR: CTAGTCTAGA

GTATTTTCTCTTTAGCTTACTG (containing XbaI site) and GCGCACTAGCGGCCGC

CACAGAACCCAATAGAAGACAGAACCC (containing NotI site).

GTATATGTGATGAGCAGAAGAGGTTATCAaCgTcAAATTGTTTTGGTTCTAAATTTGGA

ACAG and

CTGTTCCAAATTTAGAACCAAAACAATTTgAcGtTGATAACCTCTTCTGCTCATCACAT

ATAC.

EPAS1. Primers for the first PCR: GGTATAGTGACCCCGTCCAC and

TGGGTTCCCCATAGGATACA. Primers for the second PCR: CTAGTCTAGA

CTTTGCAACTCCCTGGGTAAGAG (containing XbaI site) and GCGCACTAGCGGCCGC

GTTAAGCTCCCATACAGTTATAATGTTG (containing NotI site). Mutation oligonucleotides:

GTAAATCATATTTTAGCTGCACGaCgTaACCCCACACAGGGTGGCAG (site 1) and GATTTCCTGTGTGTGTTGCCCTGaCgTaAAGGGCATTTTACCCTTGCAG (site 2)

LPIN1. Primers for the first PCR: GGATGGGCCCCTTACAGTAT and TAGTGAACAAGCGCAGAGGA. Primers for the second PCR: CTAGTCTAGA GACTGACCTTCTTTTGGAATTTCTG (containing XbaI site) and GCGCACTAGCGGCCGC CCTGCAAAAGGGAATCTGCACATGC (containing NotI site). Mutation oligonucleotides: GGTAAAAGATTTTGCTTTACTTTTCGAAaCgTaATTTTTTTAAAGAGTGTTTTACTCCAA CGATTG and CAATCGTTGGAGTAAAACACTCTTTAAAAAAATtAcGtTTCGAAAAGTAAAGCAAAAT CTTTTACC.

ELOVL6. Primers for the first PCR: AAACTGTGCGAGCACAACAC and GGCCCTTTGTTGTCAACTGT. Primers for the second PCR: CTAGTCTAGA GCTGTAGACCCCATGAGAAAAGATG (containing XbaI site) and GCGCACTAGCGGCCGC GTTATGAGTGTGTGTGAAGTCAAAC (containing NotI site). Mutation oligonucleotides: GTAATTTCTACTCTTAAGTGAGATATGAAaCgTaATCCTTTTGTTCAGTTGCCCC and GGGGCAACTGAACAAAAGGATtAcGtTTCATATCTCACTTAAGAGTAGAAATTAC.

LMBRD2. Primers for the first PCR: TGGGGGATTTTTGCATGTAT and GGTGGGATACTTGAGGGTGA. Primers for the second PCR: CTAGTCTAGA CTTCAAGTGATGCGTGTCTGTG (containing XbaI site) and GCGCACTAGCGGCCGC GGTTAAAAACAGTCTATTTAGTAAG (containing NotI site). Mutation oligonucleotides: GTTGATTATGATATTCAAGGATAAATCCTAGACGTAAATGAAAAATAGTTTTTTTTTA AGTAGTCAAAGAG and CTCTTTGACTACTTAAAAAAAAACTATTTTTCATTTACGTCTAGGATTTATCCTTGAA TATCATAATCAAC.

IL6R. Primers for the first PCR: AGAACCATGCTAGCGGAAGA and AAGACCACCAACTCCACCTG. Primers for the second PCR: CTAGTCTAGA GTGTCTCAATGCACGGGGCATTTC (containing XbaI site) and GCGCACTAGCGGCCGC CAGCCTGCCCCACTGGGCCCTG (containing NotI site). Mutation oligonucleotides: GGCCTAAAGTAAAATGATCAATAATGTTTGTAACGTAAATGAAATATTTTCAAGAAA TGTGTCCAGGGG and CCCCTGGACACATTTCTTGAAAATATTTCATTTACGTTACAAACATTATTGATCATTT TACTTTAGGCC.

Modifications to 3P-Seq protocol. The following modifications were made to the previously published method (Jan et al., 2011) to expedite the protocol and improve yield. 1) The splint ligation was done at 18°C overnight instead of 22°C; 2) RNA was phenol/chloroform extracted, passed through a P-30 column (Bio-Rad) and ethanol-precipitated after RNAse H-mediated release from streptavidin beads; 3) 1 µl of RNAse T1 was used instead of 3 µl; 4) For the 3' adapter ligation, the reaction volume was scaled to 8 µl, 12.5% PEG8000 was added, and ethanol precipitation prior to gel purification was omitted; 5) a 3:1 ethanol to water ratio instead of a 2.5:1 ratio was used for all precipitations, and 1.5 µl glycoblue (Life Technologies) was always added; 6) After both gel purifications, rather than eluting overnight, the gel slices were crushed by spinning through a 20-gauge hole in a 0.5 mL tube, frozen for 20 minutes at -80°C in 800 µl 0.3 M NaCl, thawed at 50°C for 20 minutes with occasional vortexing, and rotated at room temperature for 20 minutes. The updated protocol is available in its entirety online at bartellab.wi.mit.edu/protocols/.

Processing of 3P-seq reads and determining mRNA poly(A) sites. Using Bowtie (Langmead et al., 2009), of those tags with at least one untemplated adenylate residue, 88 million mapped

uniquely to the respective genome, including ~18 million from human, ~10 million from mouse, and ~60 million from zebrafish (Table S2A). We considered as a cluster those 3P-tags ending 31 nucleotides within the most frequently implicated cleavage position, which was then regarded as a candidate poly(A) cleavage site. To control for PCR amplification bias, we only used those clusters with at least two identifiably independent reads in our downstream analysis (e.g. tags with different ends, different number 3'-terminal adenylates, and/or from two different libraries).

28–78% of clusters (46–94% percent of tags) from each cell type mapped to 3'UTR and flanking region, which were considered as at most the 5K nucleotides downstream from the longest annotated 3'UTR end for human and mouse genes (or, in the case of zebrafish genes, 3K nucleotides) and not overlapping with the downstream gene (Table S2). 6–16% of clusters (1–8% percent of tags) mapped to upstream exons or introns of genes, and 0.1–5% of clusters (0.1–3% percent of tags) were assigned to non-coding RNAs (ncRNAs). The remainder of the clusters might be poly(A) cleavage sites for unannotated mRNAs or long non-coding RNAs (Nam and Bartel, 2012; Ulitsky et al., 2012).

Weighted 3'UTR length and weighted minimum distance of site. For each mRNA, we estimated the weighted 3'UTR length by summing the 3'UTR length of each isoform weighted by the corresponding 3'UTR isoform ratio. Similarly, the weighted minimum distance of site was estimated by summing the minimum distance of site weighted by the ratio of corresponding 3'UTR isoform embedding the site. To compare these lengths with those determined by 3'-seq (Lianoglou et al., 2013), datasets for HeLa and HEK293 cells were downloaded from http://cbio.mskcc.org/leslielab/ApA/atlas, using the denoised datasets.

Affected isoform ratio (AIR). AIR is the faction of 3'UTR isoforms containing a certain miRNA site. An AIR of 1 corresponds to a site located in the constant UTR region, whereas an AIR less than 1 corresponds to a site located in the variable region.

Processing of RNA-Seq reads and measuring expression level. All RNA-seq data were mapped to the human reference genome (hg19) using Bowtie version 0.12.8 (Trapnell et al., 2009), allowing at most five genomic matches but choosing the best one (-n 1 -e 240 -m 5 --best --strata). To measure expression level, we estimated both reads per kilobases per million reads (RPKM) values and reads per million reads (RPM) values based on RefSeq annotation. Quantile normalization was performed to reduce technical global bias of expressions between replicates and between wild type and transfected cells (Bolstad et al., 2003).

miRNA target response analysis. For RNA-seq data, we used an RPM cutoff of 3, which dramatically reduced the variance observed. For the precompiled 74 microarray data analysis, we only used genes with more than 10 RPM (as measured by RNA-seq for HeLa samples) due to high systematic noise. For mouse and zebrafish microarray data analysis, we only used top half highly expressed genes. To check miRNA target responses (for transfection and knockout), we performed cumulative distribution function (CDF) analyses and Sylamer analyses (van Dongen et al., 2008) (Figure S1A, B). For data with replicates, we also performed those analyses using mean values of replicates.

Human context score models of TargetScan. TargetScan 5.2 predicts miRNA targets using a context score model (context-only), a simple linear regression model (using least squares regression) that estimates expected $\log_2$ expression fold changes (efficacy) by taking account of site types (8mer, 7m8, and 7A1), a site location from both 3'UTR ends (or called min distance), local A/U within 30 nt upstream and downstream of site, and 3' supplementary pairing (Grimson et al., 2007). The parameters of the model were estimated using microarray data of 11 human miRNA transfection to HeLa. The model calculates a context score for each site and estimates the expected response of a target by a miRNA by summing context scores (called total context score)

for all sites on the target. The recently updated TargetScan 6.2 uses a new model (context+), which incorporates seed pairing stability (SPS) and target abundance (TA) as additional context features to consider different efficacies of miRNAs (Garcia et al., 2011). The parameters of context+ model were estimated using microarray data of 11 human miRNA transfections and 63 precompiled microarray data of miRNA/siRNA transfection. For benchmarking models, we used parameters estimated using 11 microarray data. The detailed procedure for context score and total context score calculation given the models was used as described in our previous study (Garcia et al., 2011), with a minor modification that uses a upper bounded context score based on site types. An 8mer site has an upper-bound of –0.03, a 7mer-m8 site of –0.02, and a 7mer-A1 site of –0.01.

Mouse and zebrafish context score models. All context and context+ score models for mouse and zebrafish used the parameters estimated from microarray data of human 11 miRNA transfection data (Garcia et al., 2011). Species-specific TA values were estimated using the same method described previously (Garcia et al., 2011). For wContext+ score, a weighted minimum distance was calculated with corresponding 3P-seq data instead of the original value.

wContext+ score calculation. For each site, a wContext+ score was calculated by weighting a context+ score with a cell-type specific AIR. Since the context score represented a $\log_2$-scaled fold changes, the calculation was done in the natural value space and switched back to $\log_2$-scaled space (Figure 3A). For each target, we calculated total wContext+ scores by summing all wContext+ scores.

Benchmarking of wContext+ score. To assess the wContext+ score model, we benchmarked context+ with different 3'UTR annotations, including RefSeq annotation, 3P-seq-based longest annotation, and 3P-seq-based major annotation. To compare the performance of target prediction,

we used the coefficients of determinants (r^2) of models on $\log_2$-scaled expression fold changes of targets.

Determining differentially repressed targets. All expression values of genes for each miRNA transfection to cell lines were adjusted by quantile normalization. To correct global biases such as transfection bias, we performed median-based centering of $\log_2$ changes for corresponding two cell types was performed and adjusted the slope of the linear regression function of the fold changes between each pair of cell types to be 1 (Figure S1C). The mean difference of $\log_2$ fold change, $d(i)$, for each gene was divided by the standard deviation of all $\log_2$ fold changes. By shuffling the $\log_2$ fold changes, the expected $d(i)$ value was estimated for each gene. The delta (Δ) value was calculated by subtracting expected $d(i)$ from observed $d(i)$, and the significantly differential targets were selected for a given Δ cutoff. FDR was calculated as the median number of falsely called genes during the permutations for the given cutoff. The analyses were done with the R package 'samr', implementing SAM analysis as described in (Tusher et al., 2001).When analyzing expression changes following miRNA transfection, genes with APA and those without displayed similar variability (Figure S2P).

SUPPLEMENTAL REFERNCES

Bolstad, B.M., Irizarry, R.A., Astrand, M., and Speed, T.P. (2003). A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics *19*, 185–193.

Garcia, D.M., Baek, D., Shin, C., Bell, G.W., Grimson, A., and Bartel, D.P. (2011). Weak seed-pairing stability and high target-site abundance decrease the proficiency of lsy-6 and other microRNAs. Nat. Struct. Mol. Biol. *18*, 1139–1146.

Giraldez, A.J., Mishima, Y., Rihel, J., Grocock, R.J., van Dongen, S., Inoue, K., Enright, A.J., and Schier, A.F. (2006). Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs. Science *312*, 75–79.

Grimson, A., Farh, K.K.-H., Johnston, W.K., Garrett-Engele, P., Lim, L.P., and Bartel, D.P. (2007). MicroRNA targeting specificity in mammals: determinants beyond seed pairing. Mol. Cell *27*, 91–105.

Guo, H., Ingolia, N.T., Weissman, J.S., and Bartel, D.P. (2010). Mammalian microRNAs predominantly act to decrease target mRNA levels. Nature *466*, 835–840.

Jan, C.H., Friedman, R.C., Ruby, J.G., and Bartel, D.P. (2011). Formation, regulation and evolution of Caenorhabditis elegans 3'UTRs. Nature *469*, 97–101.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. *10*, R25.

Lianoglou, S., Garg, V., Yang, J.L., Leslie, C.S., and Mayr, C. (2013). Ubiquitously transcribed genes use alternative polyadenylation to achieve tissue-specific expression. Genes Dev. *27*, 2380–2396.

Nam, J.-W., and Bartel, D.P. (2012). Long noncoding RNAs in C. elegans. Genome Res. *22*, 2529–2540.

Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics *25*, 1105–1111.

Tusher, V.G., Tibshirani, R., and Chu, G. (2001). Significance analysis of microarrays applied to the ionizing radiation response. Proc. Natl. Acad. Sci. U.S.a. *98*, 5116–5121.

Ulitsky, I., Shkumatava, A., Jan, C.H., Subtelny, A.O., Koppstein, D., Bell, G.W., Sive, H., and Bartel, D.P. (2012). Extensive alternative polyadenylation during zebrafish development. Genome Res. *22*, 2054–2066.

van Dongen, S., Abreu-Goodger, C., and Enright, A.J. (2008). Detecting microRNA binding and siRNA off-target effects from expression data. Nat Meth *5*, 1023–1025.