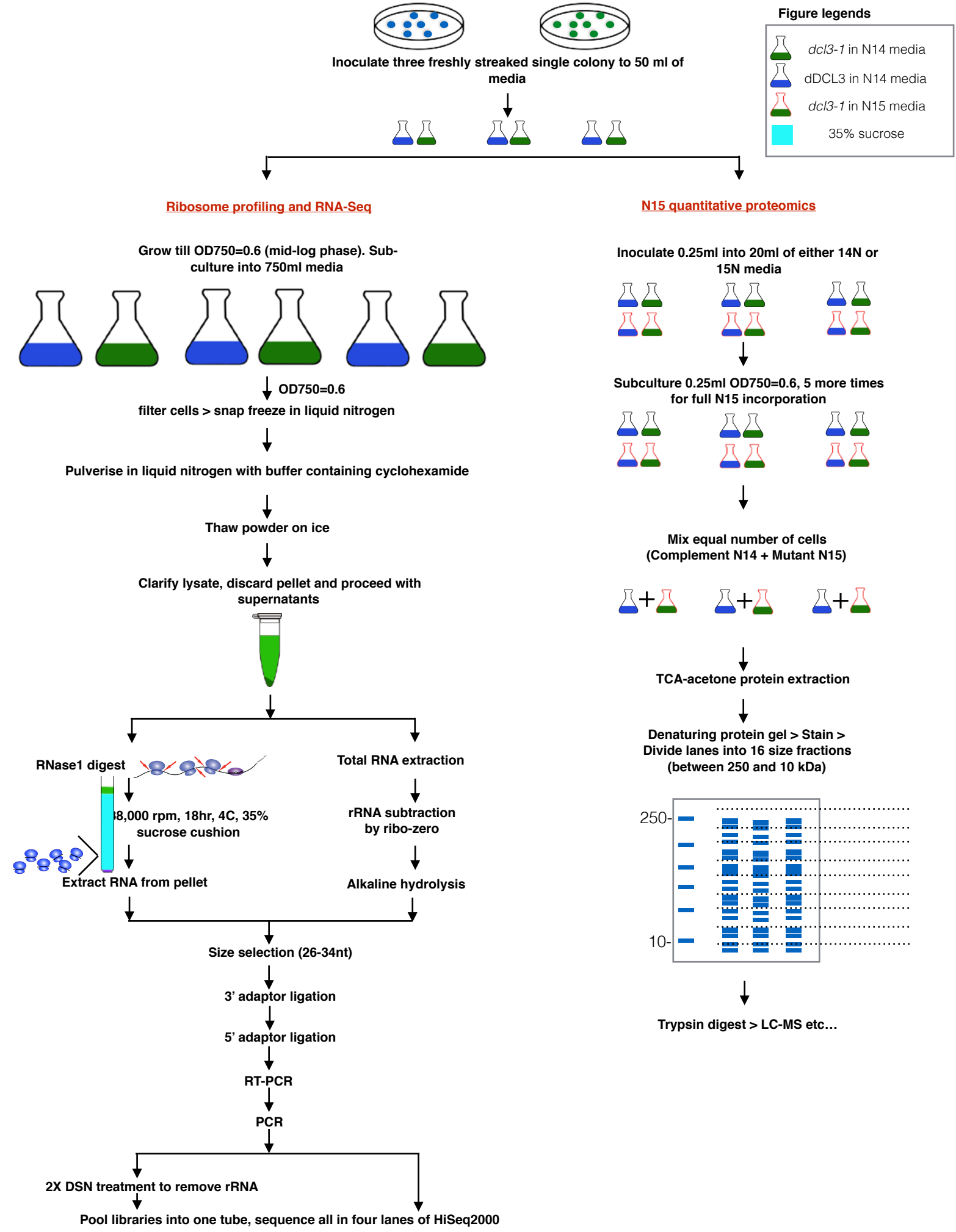
Supplementary information

Endogenous miRNA in the green alga *Chlamydomonas* regulates gene expression through CDS-targeting.

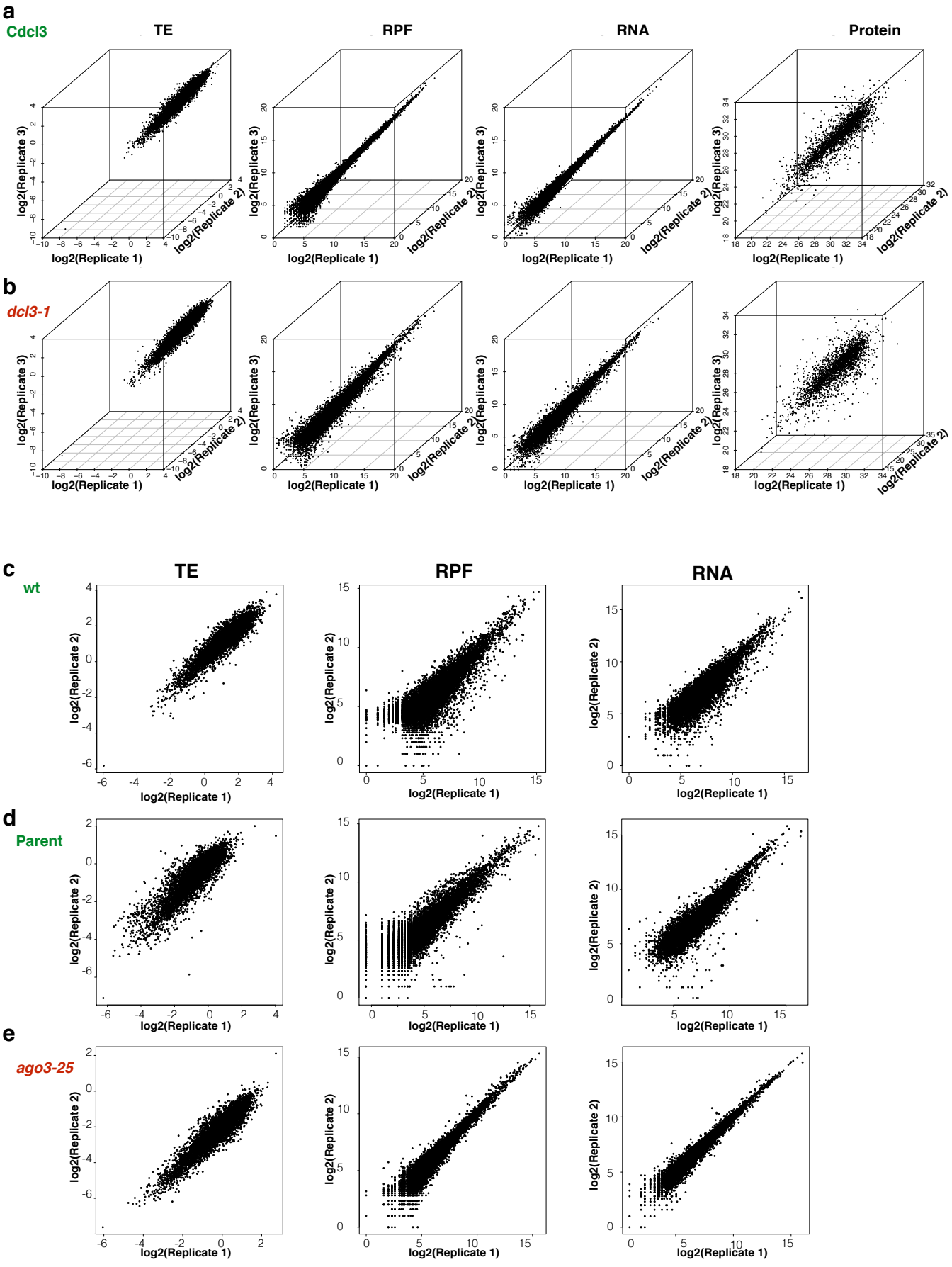
Betty Y-W. Chung*[1], Michael J. Deery[2], Arnoud J. Groen[2], Julie Howard[2] and David Baulcombe*[1].

*** Corresponding Authors: bcy23@cam.ac.uk, dcb40@cam.ac.uk**
[1] Department of Plant Sciences, University of Cambridge, Cambridge, CB2 3EA, United Kingdom
[2] Cambridge System Biology Centre and Department of Biochemistry, University of Cambridge, CB2 1GA, United Kingdom



Supplementary Figure 1. Experimental workflow

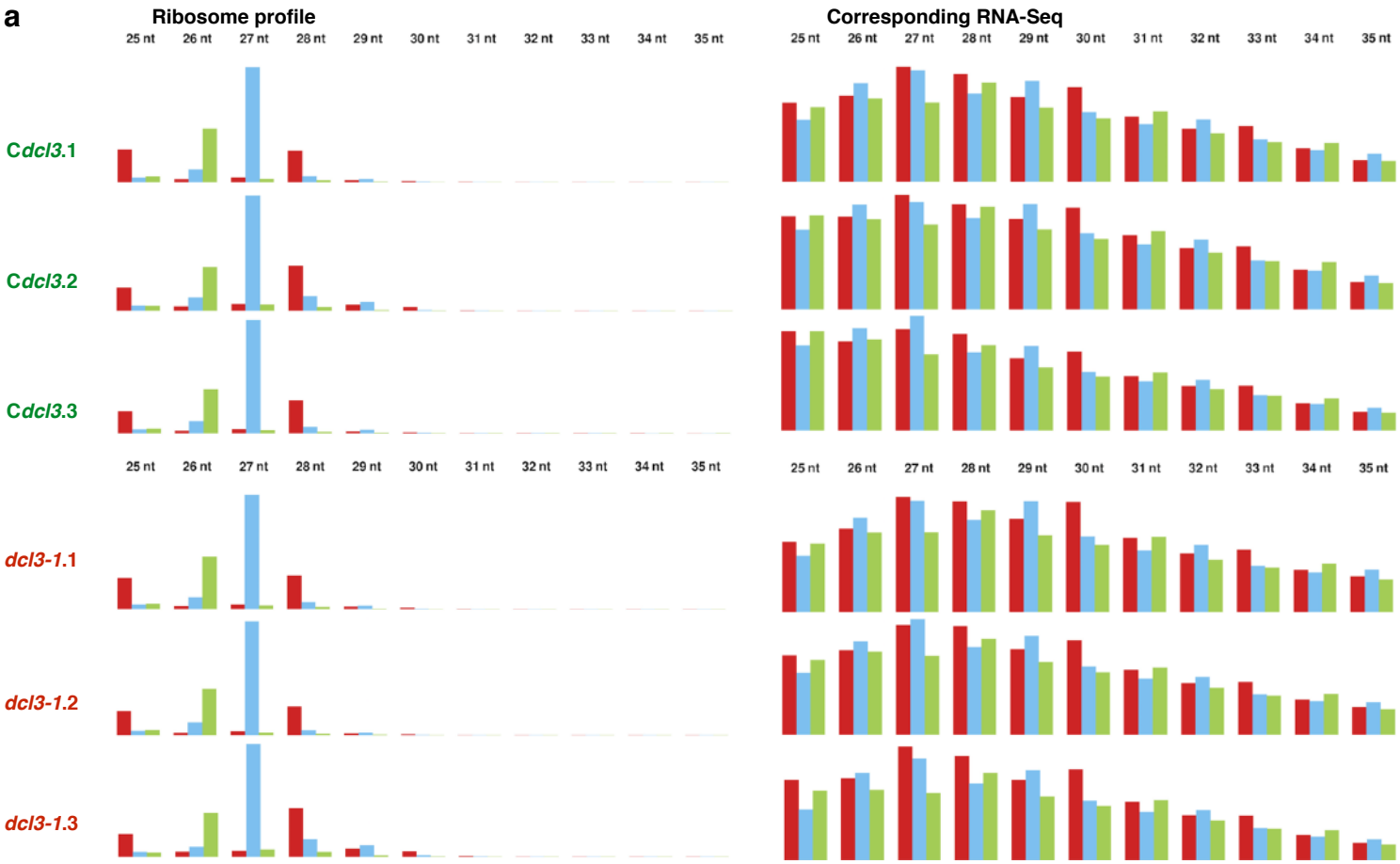
Three independent single colonies from freshly streaked *Chlamydomonas dcl3-1* (green) or complement (blue) were inoculated into 50 mL of TAP media and grown until OD750 = 0.6 (mid-log phase). 0.25 mL of each culture was used for N15 incorporation for whole cell proteomics and the remaining culture was used to sub-culture 750 mL of TAP for ribosome profiling.



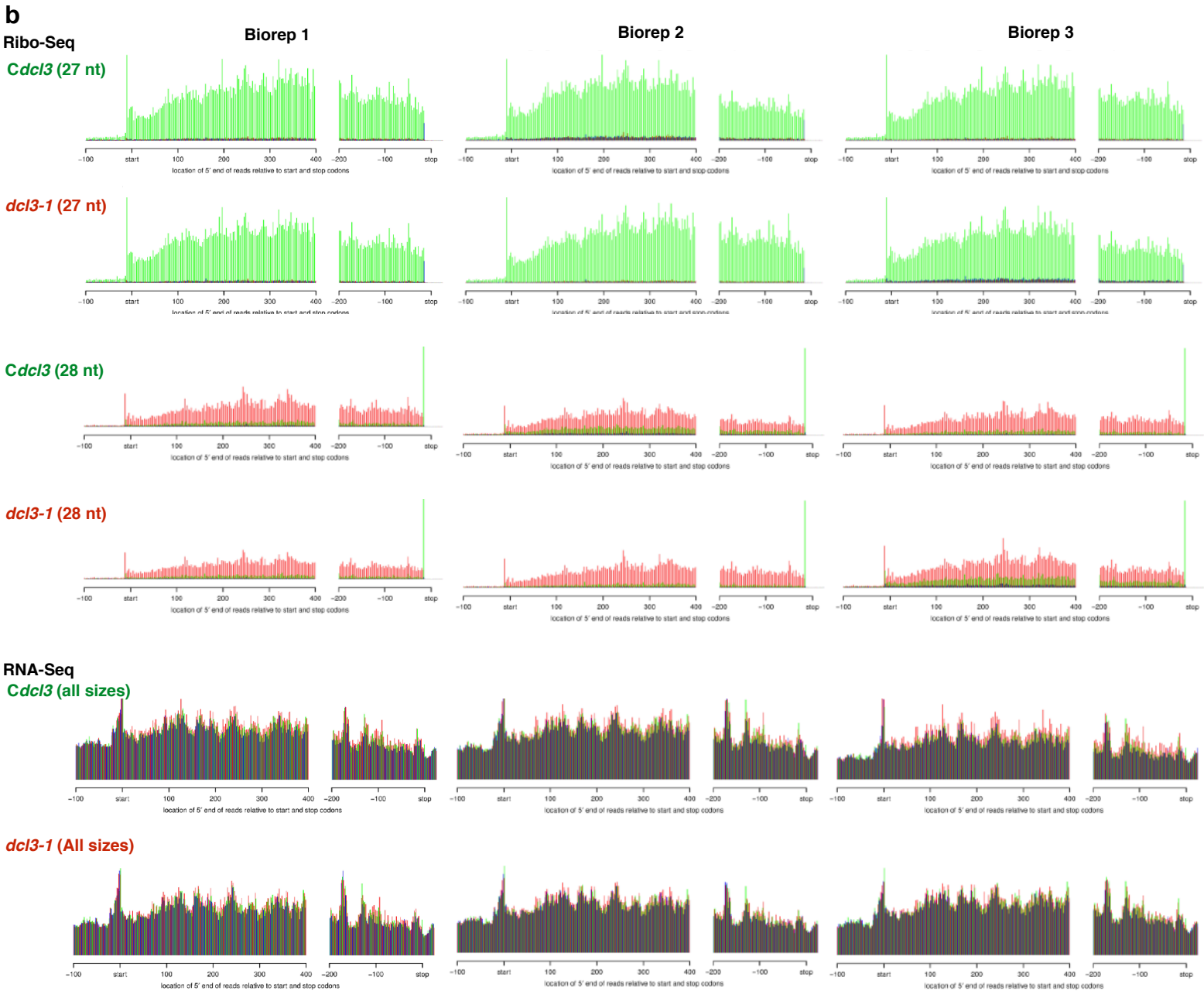
Supplementary Figure 2. Reproducibility of TE, ribosome profiling, RNA-Seq and N15 Proteomics
(A)-(E) Correspondence between biological triplicates for C*dcl3*, *dcl3-1* and replicates for wt, Parent and *ago3-25*, respectively.

Supplementary Table 1: Number of reads mapping to nuclear-encoded transcripts for each library (Phytozome 281).

		biorep 1	biorep 2	biorep 3
RiboSeq	Complement	368,613	291,373	444,768
	DCL3	515,717	461,013	590,953
RNA-seq	Complement	908,865	1,114,867	1,223,427
	DCL3	1,166,393	1,183,071	679,375

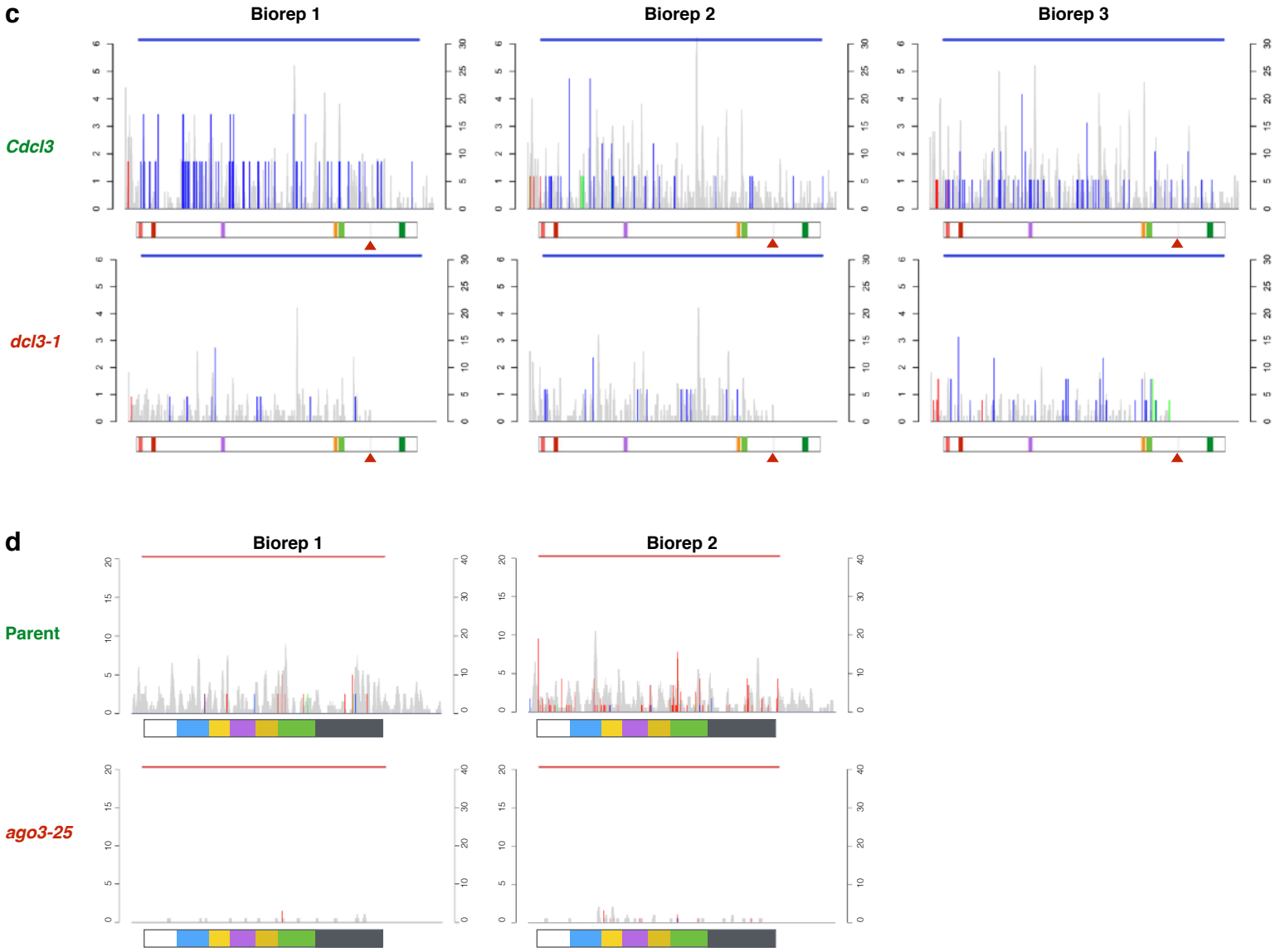


Supplementary Figure 3: Generation of precise ribosome profiling data:
(a) Histogram of positions for all biological triplicates to which the 5' ends of ribosome profile footprints (RPFs) and corresponding RNA-Seq reads map, respectively, as a function of read size class (nt), for reads mapping to the interior region of nuclear-encoded coding ORFs. Red, green and blue bars indicate the proportion of reads that map to codon positions 0, 1 and 2 (respectively).

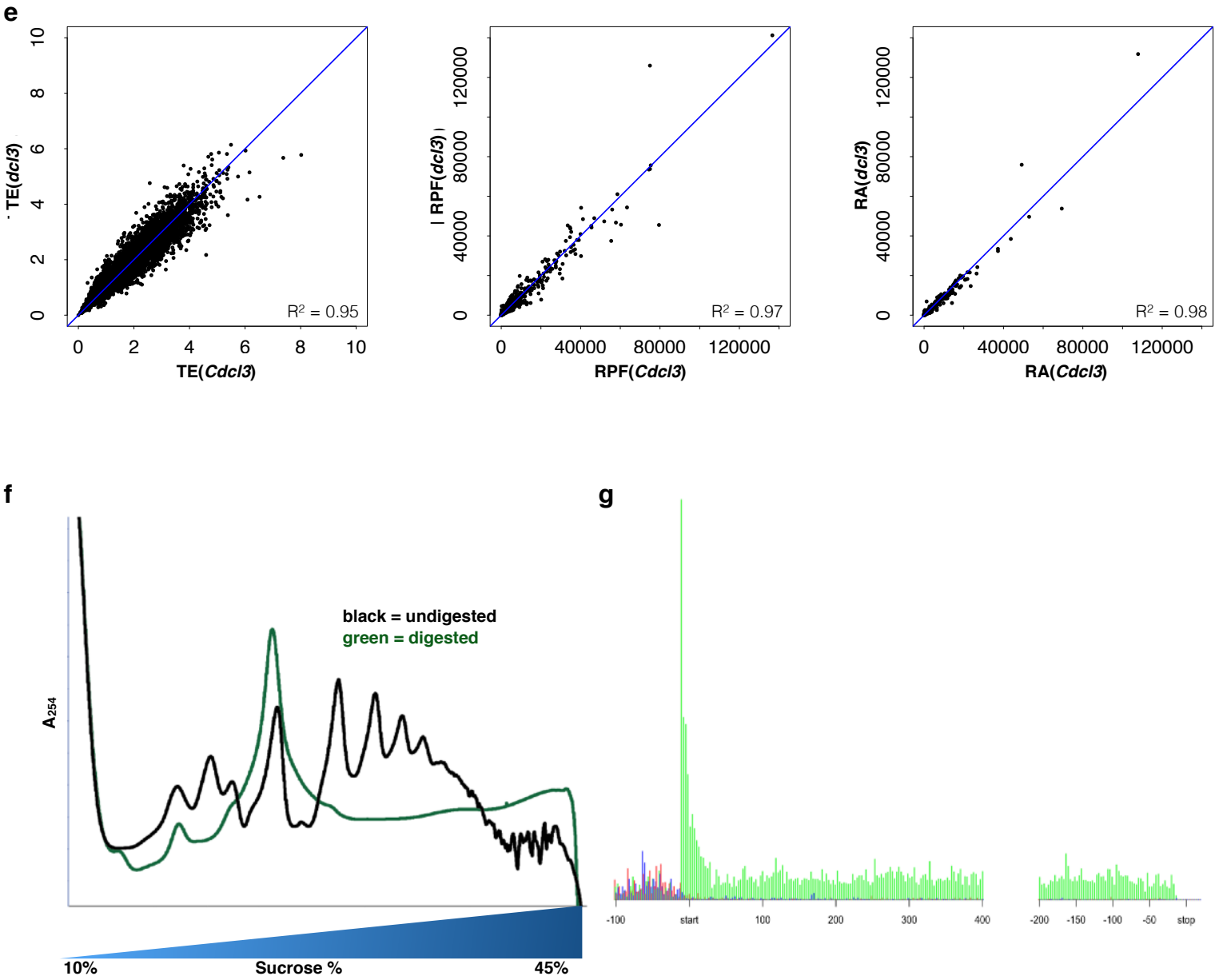


Supplementary Figure 3: Generation of precise ribosome profiling data:

(b) Histogram of 5' end positions of 27 and 28-nt RPFs and RNA-seq (all sizes) relative to start and stop codons for all biological triplicates. Reads were derived from C*dcl3* or *dcl3-1* (respectively) and summed over all transcripts. Phasing is indicated using the same colors as in supplementary figure 3A.



Supplementary Figure 3: Generation of precise ribosome profiling data:
(c) 27-nt reads mapped to DCL3 transcripts in all biological triplicates. The blue horizontal line indicates the CDS (612-12,830 nt). The schematic below the plot shows the domain organisation of DCL3 which contains two DEAD/ DEAH box helicase domains (light and dark red boxes), a Helicase domain (purple box), a proline-rich domain (orange box) and two Ribonuclease III domains a and b (light and dark green boxes, respectively). The thin grey line and the corresponding red arrow indicates the Hygromycin insertion site (nt 10,193).
(d) 27-nt reads mapped to AGO3 transcripts in all biological replicates. The red horizontal line indicates the CDS. The schematic below the plot shows the domain organisation of AGO3 which contains N-terminal domain (blue), L1 and L2 (light and dark yellow, respectively), PAZ domain (purple), MID domain (green) and the PIWI domain (grey) (Chung *et. al.* 2017 submitted).



Supplementary Figure 3: Generation of precise ribosome profiling data:

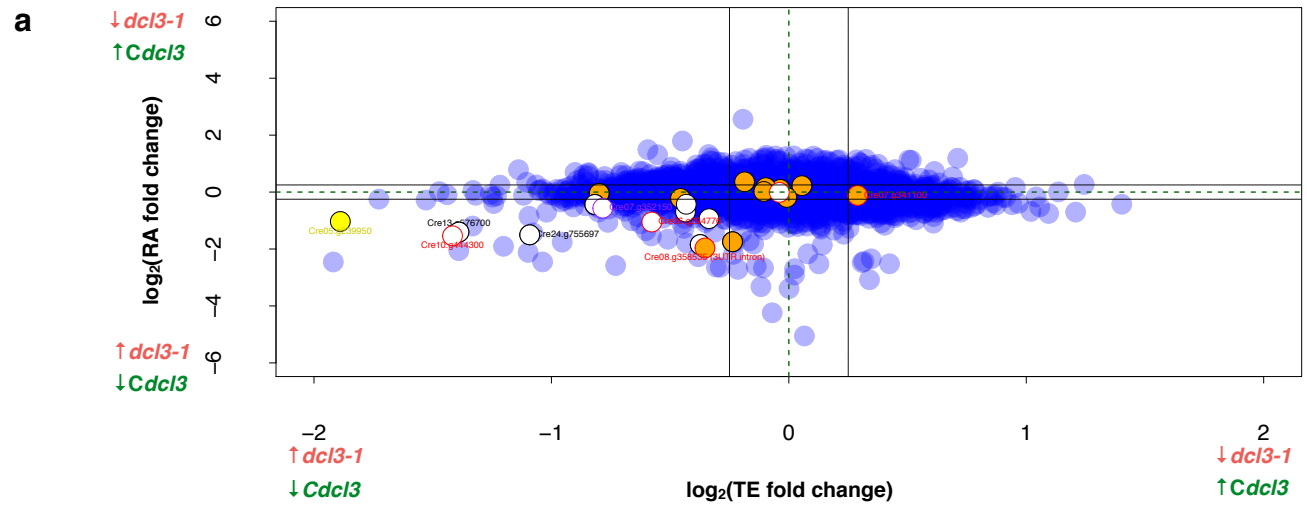
(e) Correlation of TE, RPF and RNA (averaged over biological repeats) between *dcl3-1* and C*dcl3* for all expressed genes. Blue lines represent a perfect correlation. Spearman correlation coefficients are indicated in bottom right corners.

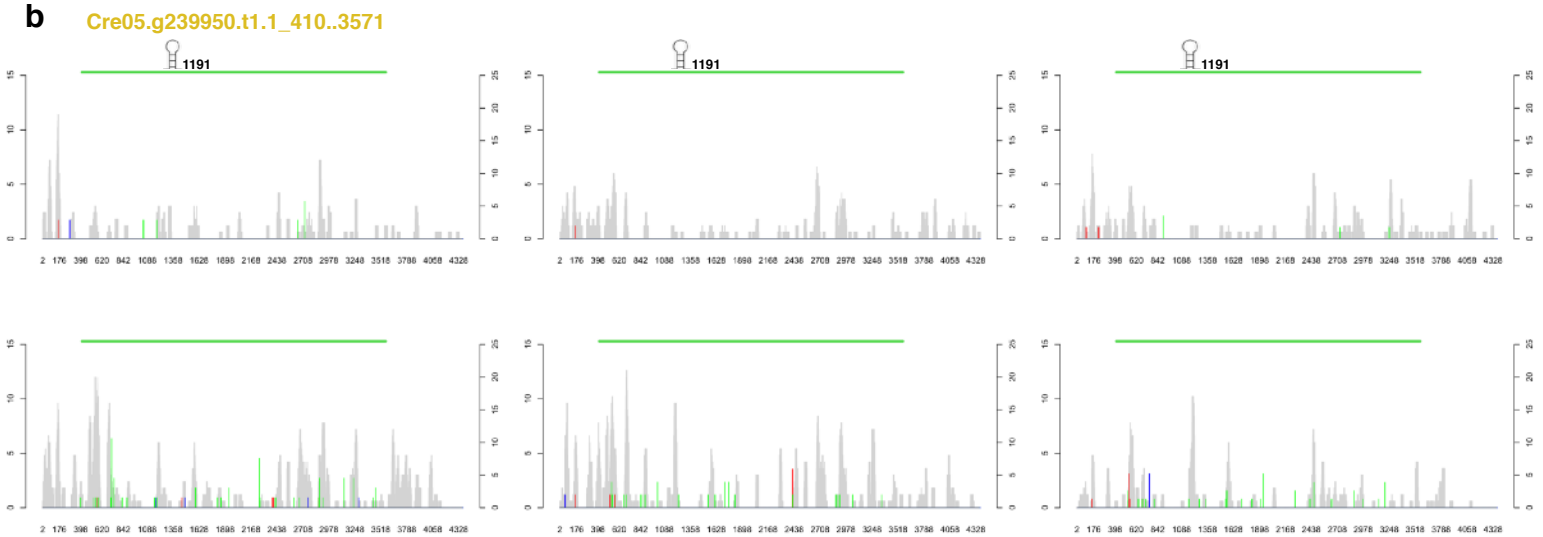
(f) Typical polysome profile of undigested (black) and digested (green) lysates in this study.

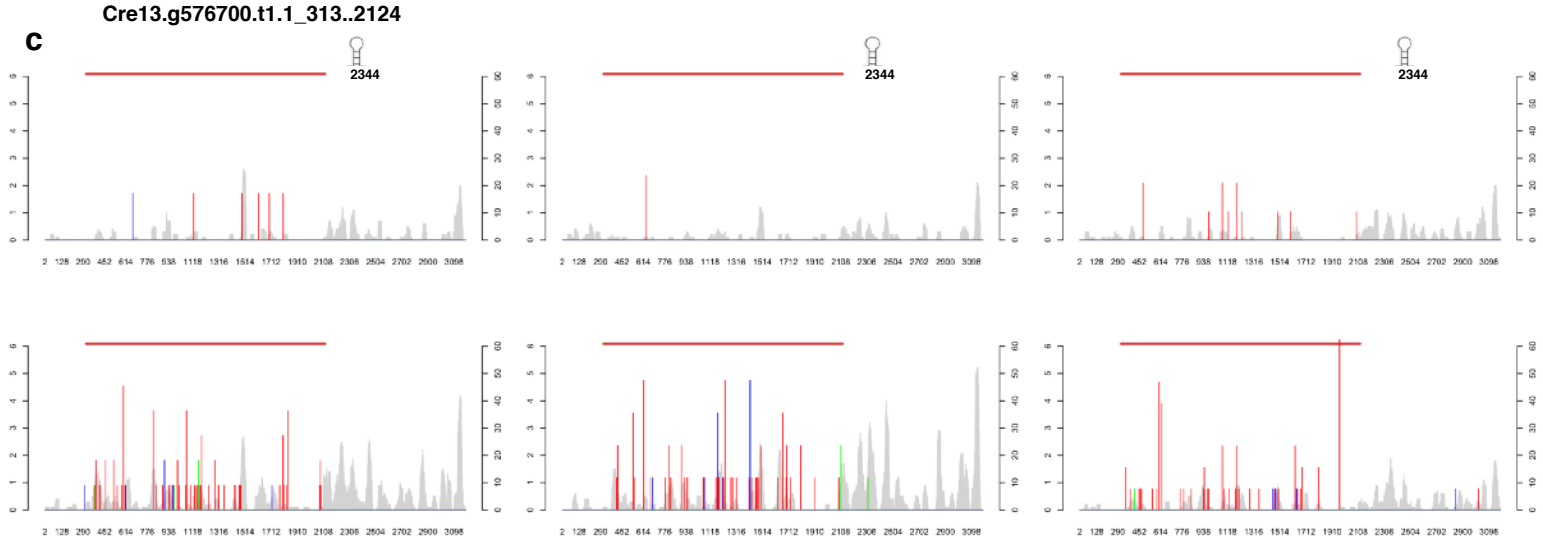
(g) Histogram of 5' end positions of 27-nt RPFs relative to start and stop codons for C*dcl3* culture pre-treated with 100 µg/ml cyclohexamide for 5 min prior to harvesting for ribosome profiling. Phasing is indicated using the same colors as in supplementary figure 3A.

Supplementary Table 2: Re-annotation of miRNA precursor-containing mRNAs.

CDS-exons	CDS-Introns	3'UTR-exon	3'UTR-intron
Cre05.g239950	Cre04.g229050	Cre14.g623850	Cre08.g358535
	Cre01.g035500	Cre02.g089850	
	Cre04.g225700	Cre02.g143427	
	Cre06.g274550	Cre02.g143527	
	Cre06.g296983	Cre03.g195950	
	Cre07.g328400	Cre05.g242301	
	Cre07.g354150	Cre10.g465000	
	Cre12.g537671	Cre13.g576700	
	Cre14.g629200	Cre13.g585175	
	Cre01.g035500	Cre13.g585200	
	Cre02.g143327	Cre16.g694950	
	Cre04.g217925	Cre02.g143527	
	Cre04.g229050	Cre24.g755697	
	Cre08.g358537	Cre02.g143527	
	Cre09.g406983	Cre13.g576700	
	Cre16.g647602	Cre16.g694950	
	Cre07.g341100	Cre24.g755697	
		Cre07.g352150	
		Cre06.g294776	
		Cre10.g444300	
		Cre01.g051050	







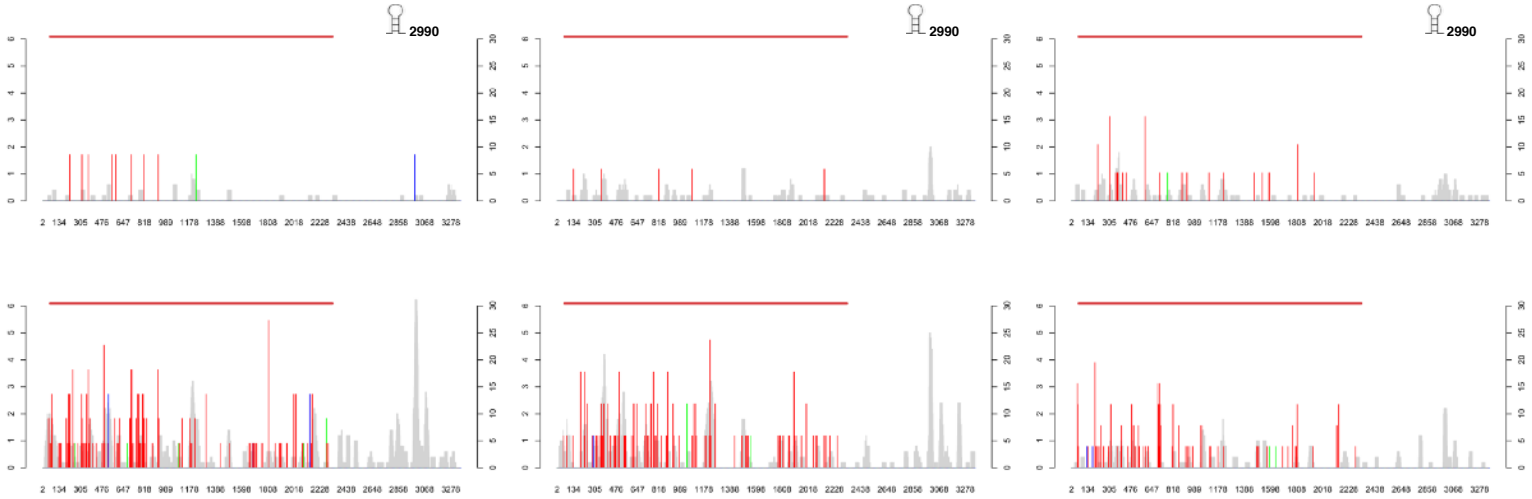
Supplementary Figure 4. DCL3-dependent processing of miRNA down-regulates translation efficiency

(a) Scatter plot of $\log_2$ fold changes of all mRNAs for TE and RA fold-changes between *dcl3-1* and C*dcl3*. New annotation for precursor-containing transcripts: yellow circle = precursor-containing CDS, white circles = precursor-containing 3'UTRs, orange circles = precursor-containing introns, white circles with red outlines = transcripts previously annotated as non-coding transcripts but which are in fact coding and contain a miRNA precursor in the 3'UTR. Pearson correlation = 0.072 and 0.486 for intron- and exon-containing transcripts respectively.

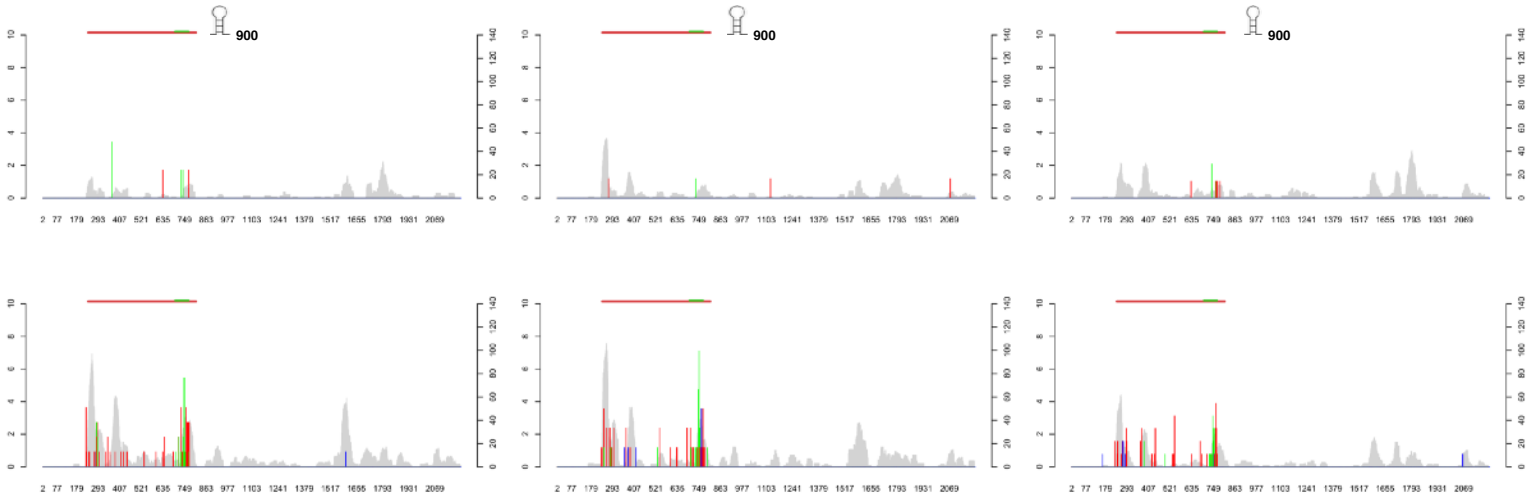
(b)-(e) Histogram of normalised 5' end positions of 27-nt RPFs relative to start and stop codons (colour) and corresponding RNA-seq reads (grey) for miRNA-precursor containing transcripts. Reads were derived from the complement or DCL3 mutant (top and bottom in biological triplicates, respectively) and summed over all transcripts.

d

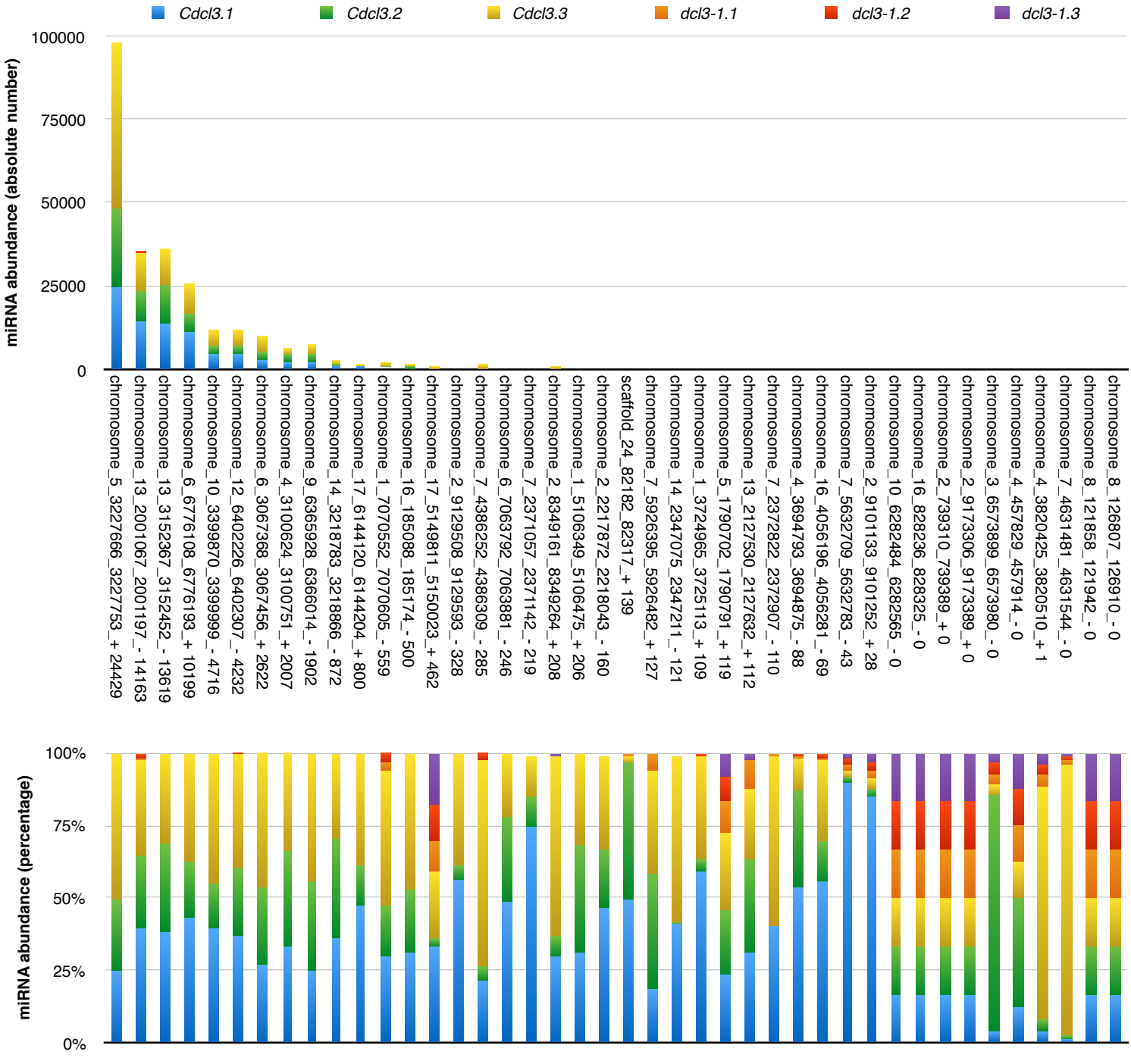
Cre24.g755697.t1.1_136..2328



e **Cre10.g444300.t1.2_241..810**



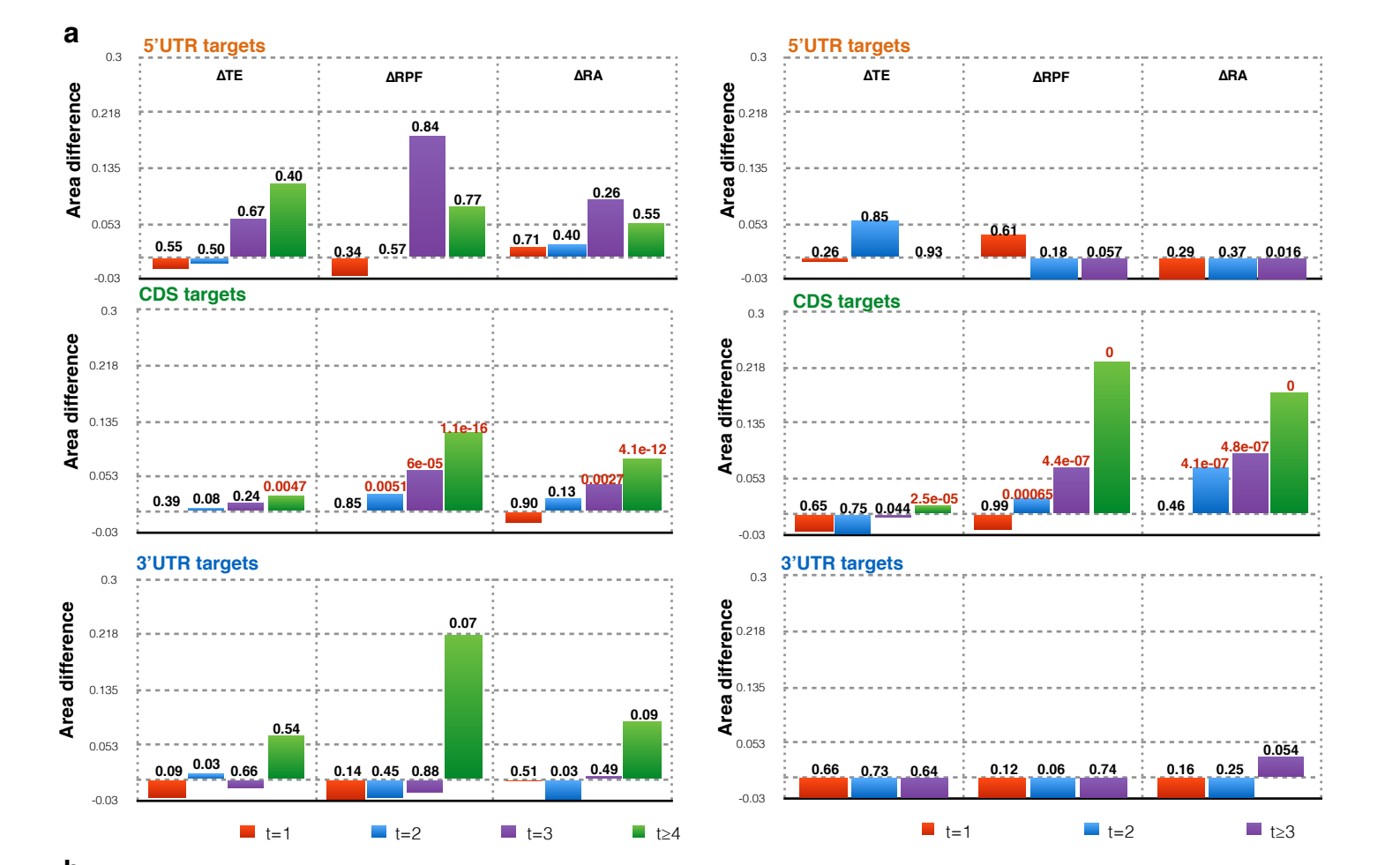
Supplementary Figure 5

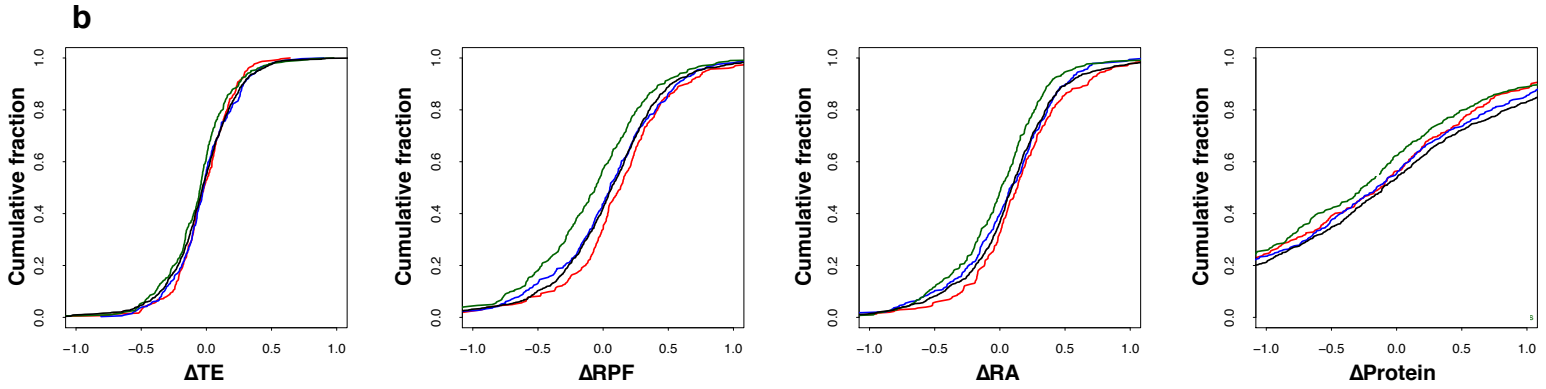


Library size	C*dcl3.1*	C*dcl3.2*	C*dcl3.3*	*dcl3-1.1*	*dcl3-1.2*	*dcl3-1.3*
trimmed reads	6,470,317	5,031,324	6,722,863	5,461,747	5,282,454	5,746,937
miRNA reads	153,598	112,892	179,859	2332	2017	2219

Supplementary Figure 5: miRNA quantification

Absolute (top) and relative (bottom) quantification for all known positive-strand miRNA reads detected in all corresponding sRNA-seq libraries. Sequencing and miRNA alignment statistics for each library are in the table below.

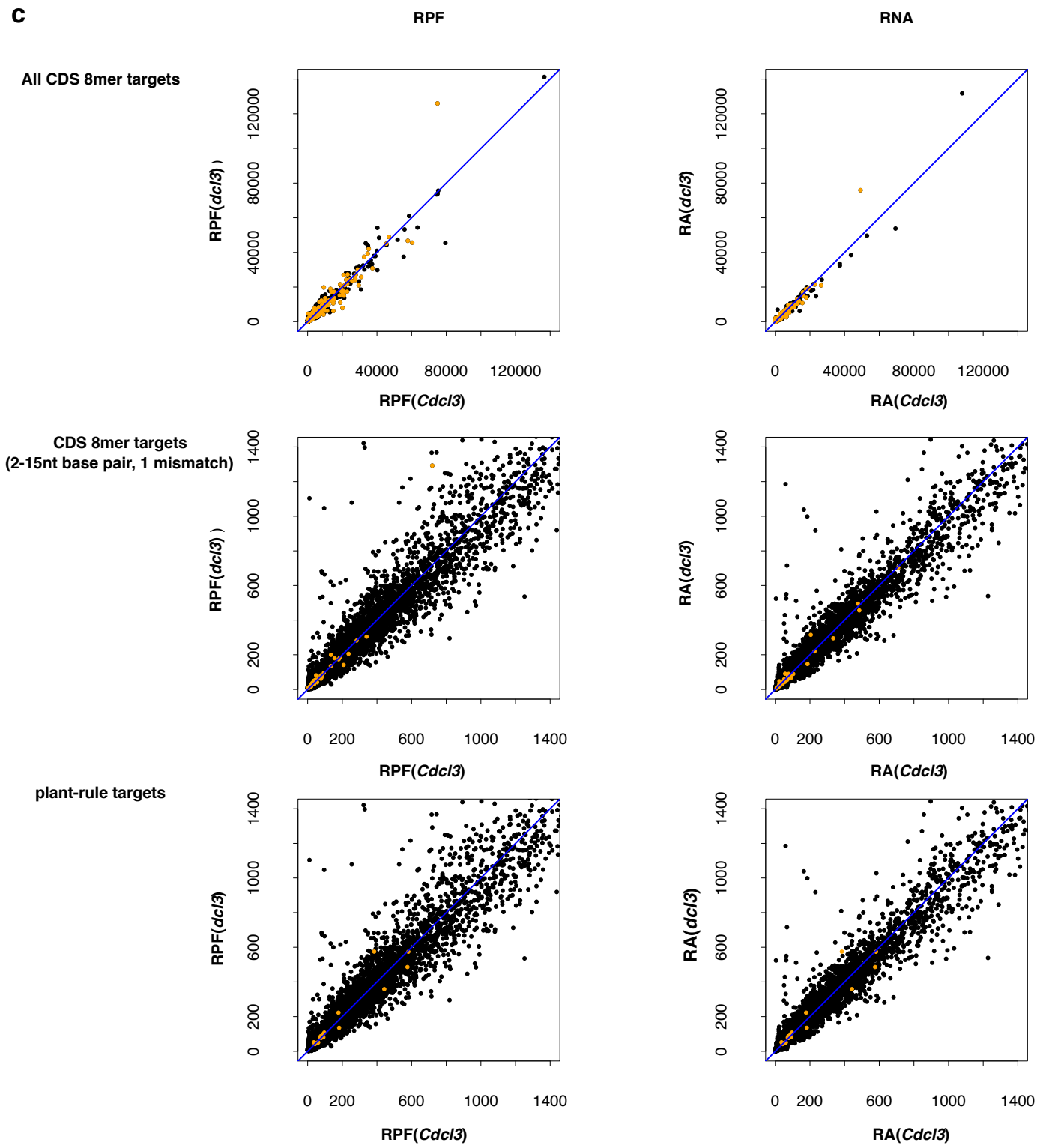




p-value with non-target (n=911) comparisons	TE	RPF	RNA	Protein
1 target site (n = 471)	0.16	0.013	0.26	0.12
2 target sites (n = 297)	0.57	0.73	0.49	0.85
≥4 target sites (n = 187)	0.05	5.3E-05	0.0026	0.043

Supplementary Figure 6:

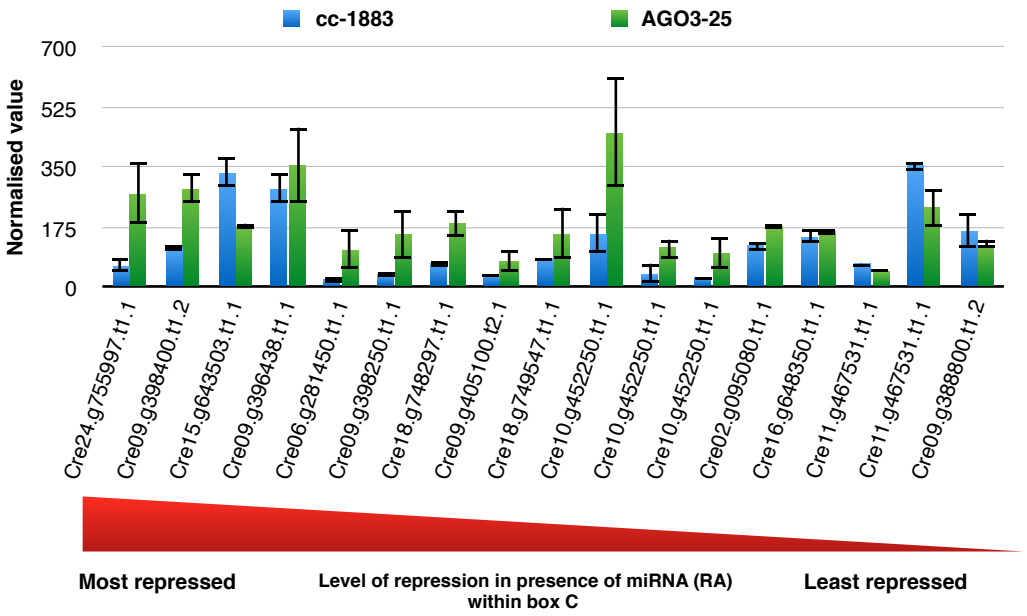
(a) Bar graph of differences between area under cumulative distribution of mRNA containing 1 (red), 2 (blue), 3 (purple) or 4 or more (green) 5'UTR, CDS or 3'UTR-exclusive target sites and non-target containing mRNAs. Significance (K.S. test) of the differences are indicated above each bar; p-values less than or equal to 0.01 are highlighted in red. Due to lower sequencing coverage, fewer mRNAs in the *ago3-25* and Parental strains passed the detection threshold, thus 5' and 3' UTR-exclusive targets containing three target sites were combined with those containing four or more target sites.
(b) Cumulative *dcl3-1* relative to C*dcl3* $\log_2$ fold-change distributions of ΔTE, ΔRPF, ΔRA and ΔProtein for genes with both NGS and proteomic support and with 0 (black), 1 (red), 2-3 (blue) or 4 or more (green) target sites. K.S. p-values are shown in the table below.



Supplementary Figure 6:

(c) Correlation of TE, RPF and RNA (averaged over biological repeats) between *dcl3-1* and C*dcl3* for all expressed mRNAs (black), and for mRNAs with predicted target sites (orange). Top, middle and bottom panels represents, respectively, all CDS-exclusive 8mer targets, CDS-exclusive targets with a higher degree of pairing (i.e. 2-15 nt base pairing, allowing 1 mismatch), and CDS-exclusive targets based on the plant rule. The blue lines represent a perfect correlation.

d



Supplementary Figure 6:

(d) Normalized mRNA expression in wild-type and *ago3-25.* Candidates are from box C of Figure 4A organized based on degree of repression in C*dcl3* relative to *dcl3-1*. Normalized average expression and standard deviation are based on biological replicates.

Supplementary table 3

Lists of mRNAs lying within boxes A, A', B, B', C and C' (Figure 4A) and their respective annotations. Annotations associated with the 80S translation machinery are highlighted in green, and other RNA binding proteins in red. Messenger RNAs with detectable protein in the N15 proteomics data are highlighted in blue.

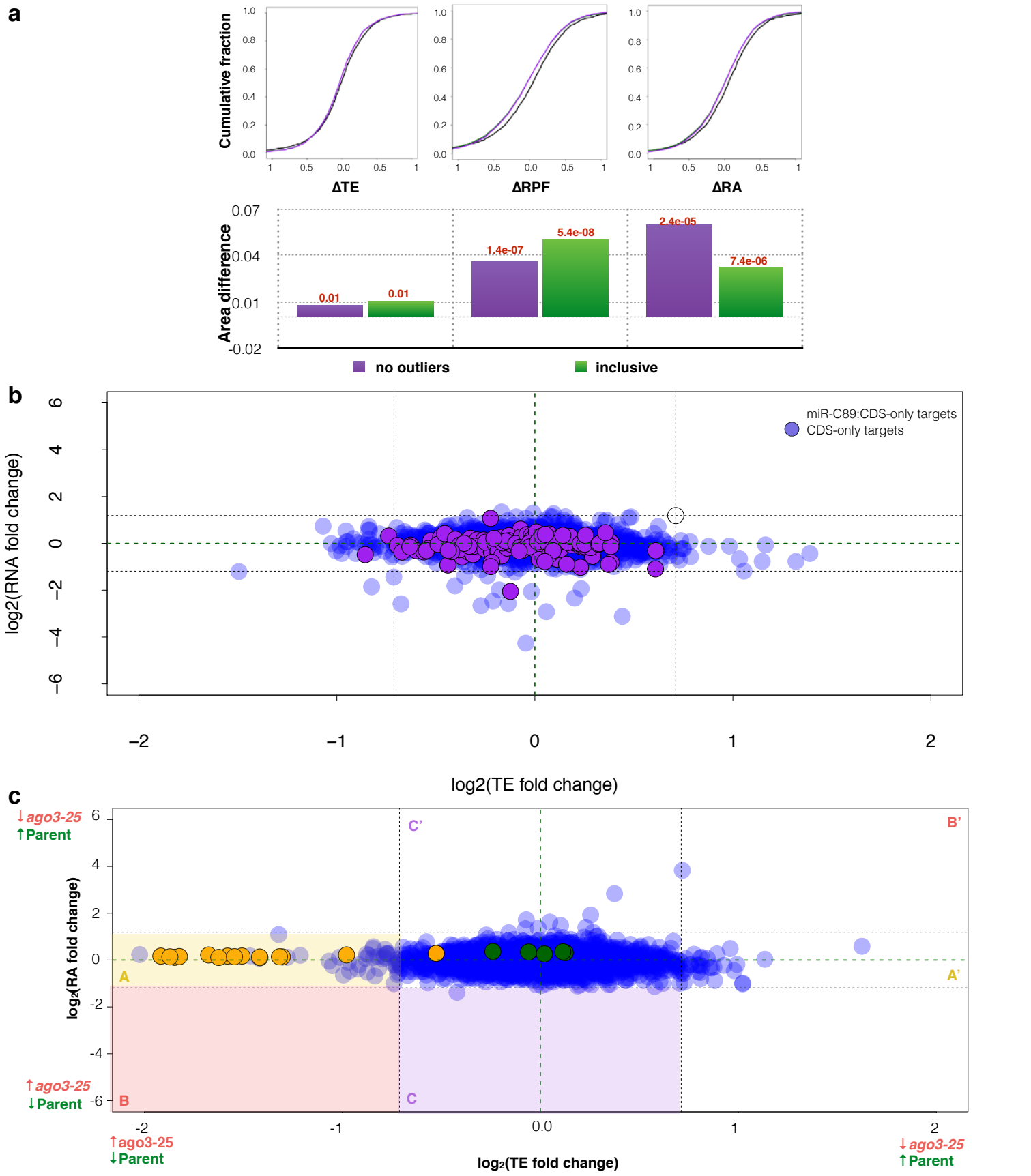
A (32)	
Cre01.g034600.t1.2	n/a
Cre03.g165215.t1.1	ubiquitin-like modifier-activating enzyme ATG7
Cre04.g222700.t1.2	Elongation factor 3
Cre09.g408950.t1.1	Autophagy-specific Gee 2, isoform A
Cre09.g397450.t1.1	Dimethylaniline Mnooxygenase
Cre12.g504200.t1.2	Ribosomal protein S23, component of cytosolic 80S ribosome and 40S small subunit
Cre17.g701200.t1.2	Ribosomal protein L14, component of cytosolic 80S ribosome and 60S large subunit
Cre08.g359450.t1.2	D-Alanine Ligase
Cre02.g093850.t1.1	Ras supressor protein (contains leucine-rich repeats)
Cre11.g468353.t1.1	SF14-voltage-gated potassium channel
Cre04.g214503.t1.1	Ribosomal protein S12, component of cytosolic 80S ribosome and 40S small subunit
Cre02.g091100.t1.2	Ribosomal protein L15, component of cytosolic 80S ribosome and 60S large subunit
Cre02.g106600.t1.2	Ribosomal protein S19, component of cytosolic 80S ribosome and 40S small subunit
Cre12.g498900.t1.2	Ribosomal protein S7, component of cytosolic 80S ribosome and 40S small subunit
Cre06.g299450.t1.2	n/a
Cre06.g280800.t1.2	Nuclear auto antigenic sperm protein
Cre07.g349950.t1.1	Transcription initiation factor RFIID subunit 6
Cre12.g521200.t1.2	DNA replication factor C complex subunit 1
Cre17.g720300.t1.2	Non-specific serine/threonine protein kinase
Cre06.g272950.t1.1	Ribosomal protein S18, component of cytosolic 80S ribosome and 40S small subunit
Cre08.g385800.t1.1	n/a
Cre03.g174900.t1.1	SARM1 (protein binding)
Cre12.g494050.t1.2	Ribosomal protein L9, component of cytosolic 80S ribosome and 60S large subunit
Cre09.g399141.t1.1	MFS transporter, ACS family, solute carrier family 17
Cre11.g467560.t1.1	TPR repeat containing protein
Cre16.g660750.t1.1	coiled-coil and C2 domain-containing protein 2A
Cre07.g357850.t1.2	Ribosomal protein L22, component of cytosolic 80S ribosome and 60S large subunit
Cre01.g040850.t1.2	G Protein-coupled receptor-related protein
Cre01.g036800.t1.1	Diacylglycerol kinase
Cre16.g661588.t1.1	FAST Leu-Rich Domain-containing protein
Cre07.g348550.t1.1	CGI-141-related/Lipase containing protin
Cre01.g023550.t1.1	Flagellar Associated Protein, putative outer arm dynein light chain

B (3)	
Cre16.g675200.t1.1	n/a
Cre12.g541400.t1.2	Las17-binding protein actin regulator (Ysc84)
Cre16.g677920.t1.2	n/a

C (15)	
Cre02.g095080.t1.1	Major vault protein
Cre03.g155950.t1.2	n/a
Cre06.g281450.t1.1	Scavenger receptor cysteine rich (SRCR) protein
Cre09.g388800.t1.2	Glutamate dehydrogenase
Cre09.g396438.t1.1	n/a
Cre09.g398250.t1.1	n/a
Cre09.g398400.t1.2	Transient receptor potential ion channel protein
Cre09.g405100.t2.1	n/a
Cre10.g452250.t1.1	n/a
Cre11.g467531.t1.1	Flagellar Associated Protein
Cre15.g643503.t1.1	n/a
Cre16.g648350.t1.1	Proline Oxidase
Cre18.g748297.t1.1	n/a
Cre18.g749547.t1.1	n/a
Cre24.g755997.t1.1	Cell wall protein pherophorin-C18

A' (16)	
Cre16.g668050.t1.1	Aspartyl protease (Asp_protease_2)
Cre06.g265850.t1.1	Tail-specific/C-terminal processing peptidase protease
Cre03.g191950.t1.2	RimM N-terminal domain (RimM)
Cre14.g614950.t1.2	Putative mitochondrial ribosomal protein S2, imported to mitochondria
Cre12.g554300.t1.1	Sodium:solute symporter
Cre15.g639050.t1.1	Zinc finger MYND domain containing protein 10
Cre14.g626800.t1.1	n/a
Cre14.g610663.t1.1	n/a
Cre06.g271950.t1.2	General vesicular transport factor P115
Cre07.g318300.t1.1	CAMP-dependent protein kinase regulatory chain
Cre17.g739850.t1.2	n/a
Cre10.g429200.t1.1	RuBisCO methyltransferase
Cre17.g703450.t1.1	n/a
Cre14.g623439.t1.1	Pyroglutamyl-peptidase I
Cre14.g626800.t1.1	n/a
Cre14.g622150.t1.1	n/a

C' (3)	n/a
Cre43.g760497.t1.1	n/a
Cre17.g734200.t1.2	L,L-diaminopimelate aminotransferase
Cre17.g734100.t1.2	n/a

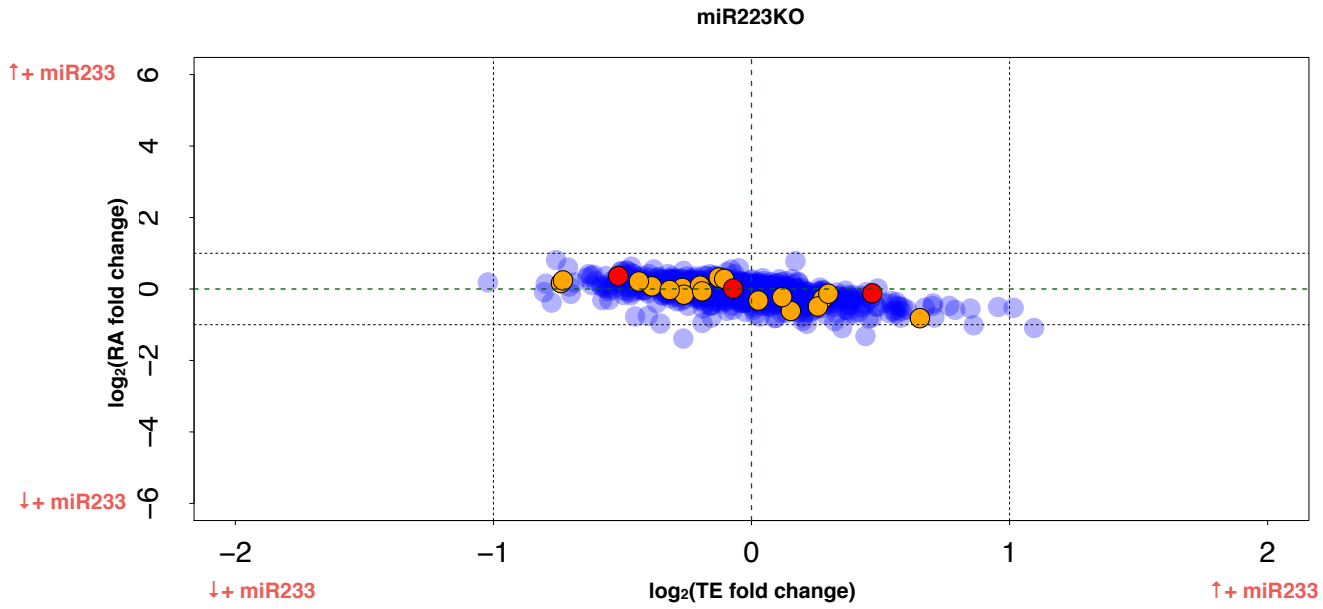


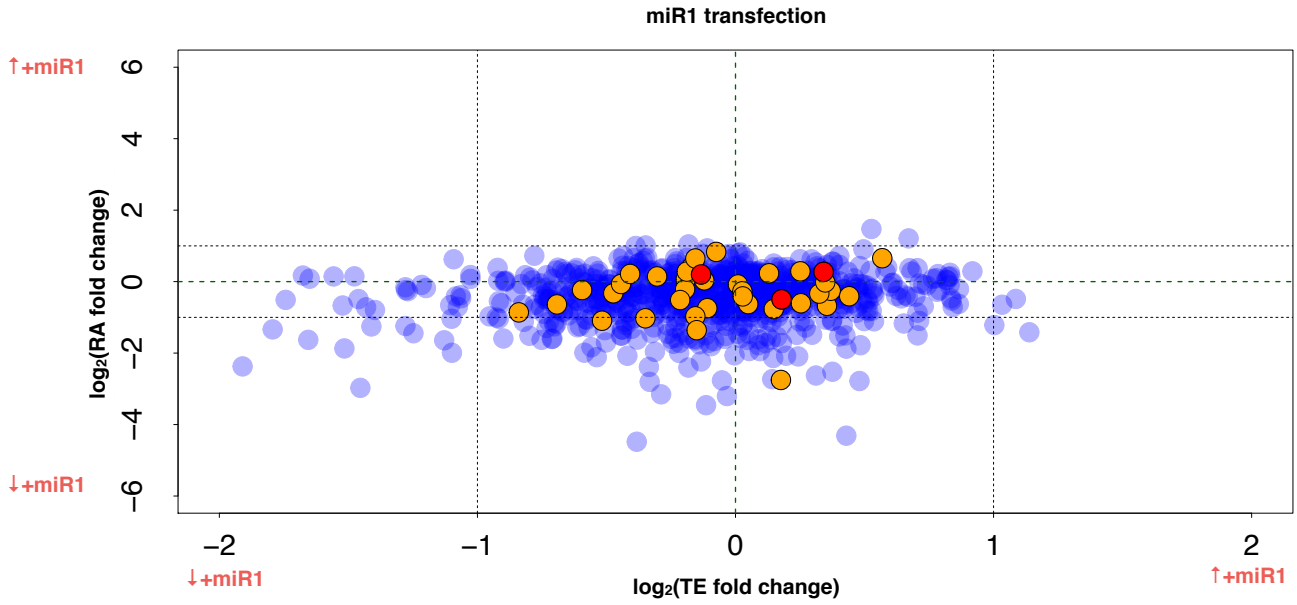
Supplementary Figure 7

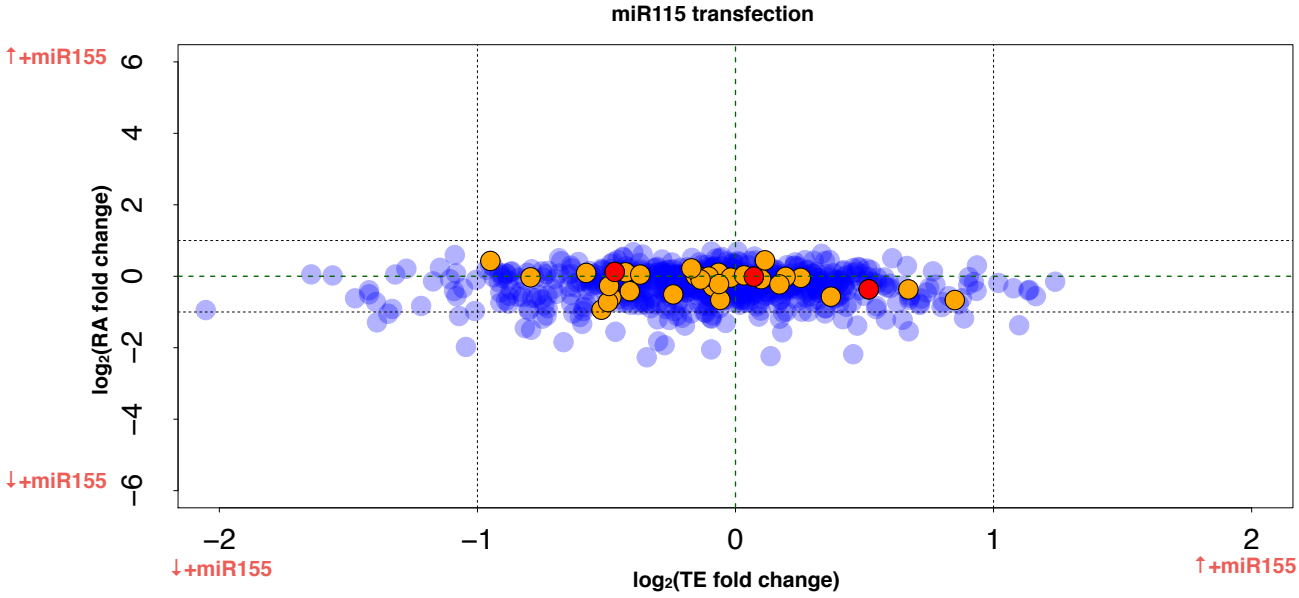
(a) Top panels show cumulative distributions of ΔTE (left), ΔRPF (middle) and ΔRA (right) $\log_2$ fold changes for CDS-exclusive targets in *dcl3-1* relative to C*dcl3*. Colors indicate CDS-exclusive targets (green), CDS-exclusive targets excluding strongly differentially expressed mRNAs (i.e. mRNAs in box A, A', B, B', C and C' in Figure 4A) (purple), and mRNAs without target sites (black). The bar graphs in the bottom panel show differences between areas under the corresponding cumulative distributions of target-site containing mRNA and non targets. Significance (K.S. test) of the differences are indicated above each bar; p-values less than or equal to 0.01 are highlighted in red.

(b) Scatter plot of $\log_2$ fold changes of all CDS-exclusive targets for ΔTE and ΔRA between *dcl3-1* and C*dcl3*. miR-C89 CDS-exclusive targets are highlighted in purple.

(c) Scatter plot of $\log_2$ fold changes of all CDS-exclusive targets for ΔTE and ΔRA between *ago3-25* and the Parental strain. 80S and chloroplast ribosomal proteins are in orange and green, respectively.







Supplementary Figure 8

Correspondence between ΔTE and ΔRA $\log_2$ fold-changes after deleting miR-233 in mouse neutrophil cells (A), after introducing miR-1 to HEK293 cells (B), and after introducing miR-155 to HEK293 cells (C). Fold-change data were obtained from[9].