SUPPLEMENTAL INFORMATION

Gene- and species-specific Hox mRNA translation by ribosome expansion segments

Kathrin Leppek[1,2], Kotaro Fujii[1,2], Nick Quade[3], Teodorus Theo Susanto[1,2], Daniel Boehringer[3], Tea Lenarčič[3], Shifeng Xue[1,2,†], Naomi R. Genuth[1,2], Nenad Ban[3*], and Maria Barna[1,2,4,*]

[1] Department of Developmental Biology, Stanford University, Stanford, California 94305, USA
[2] Department of Genetics, Stanford University, Stanford, California 94305, USA
[3] Department of Biology, Institute of Molecular Biology and Biophysics, Otto-Stern-Weg 5, ETH Zürich, Zürich 8093, Switzerland
[†] present address: Department of Biological Sciences, National University of Singapore, 14 Science Drive 4, Singapore 117543
[4] Lead Contact: Maria Barna, mbarna@stanford.edu
* co-corresponding authors, contact: ban@mol.biol.ethz.ch; mbarna@stanford.edu

This document includes:

Figures S1 to S7
Tables S1 to S3

Other supplementary material for this manuscript includes the following:

Table S4: Relative protein quantification with TMT labelling for 4xS1m pulldown from C3H/10T1/2 cells
Table S5: Relative protein quantification with TMT labelling for 4xS1m pulldown from mouse embryos

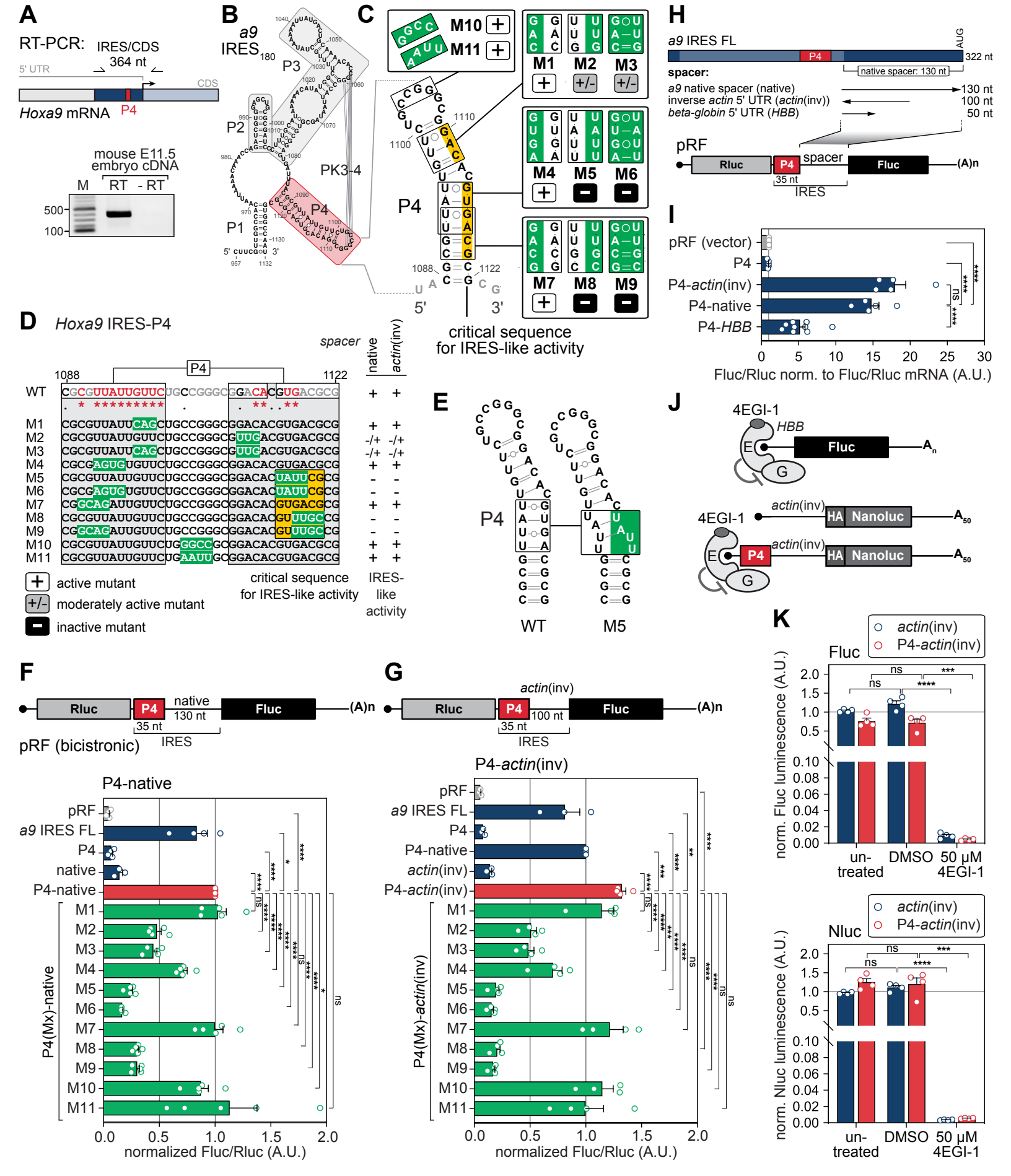
A
RT-PCR:
IRES/CDS
364 nt
5' UTR
CDS
Hoxa9 mRNA
P4
mouse E11.5
embryo cDNA
M
RT
- RT
500
100
B
a9
IRES
180
P3
P2
PK3-4
P4
P1
5' CUUCGCU 3'
C
M10
M11
M1
M2
M3
M4
M5
M6
M7
M8
M9
P4
5'
3'
critical sequence
for IRES-like activity
H
a9 IRES FL
P4
AUG
322 nt
native spacer: 130 nt
spacer:
a9 native spacer (native)
130 nt
inverse actin 5' UTR (actin(inv))
100 nt
beta-globin 5' UTR (HBB)
50 nt
pRF
Rluc
P4
spacer
Fluc
(A)n
35 nt
IRES
I
pRF (vector)
P4
P4-actin(inv)
P4-native
P4-HBB
ns
****
Fluc/Rluc norm. to Fluc/Rluc mRNA (A.U.)
D
Hoxa9 IRES-P4
spacer
native
actin(inv)
P4
1088
1122
WT
M1
M2
M3
M4
M5
M6
M7
M8
M9
M10
M11
active mutant
moderately active mutant
inactive mutant
critical sequence
for IRES-like activity
IRES-
like
activity
E
P4
WT
M5
J
4EGI-1
HBB
Fluc
A_n
actin(inv)
HA
Nanoluc
A_50
P4
K
actin(inv)
P4-actin(inv)
Fluc
norm. Fluc luminescence (A.U.)
un-
treated
DMSO
50 μM
4EGI-1
Nluc
norm. Nluc luminescence (A.U.)
F
Rluc
P4
native
130 nt
Fluc
(A)n
35 nt
IRES
pRF (bicistronic)
P4-native
pRF
a9 IRES FL
P4
native
P4-native
M1
M2
M3
M4
M5
M6
M7
M8
M9
M10
M11
P4(Mx)-native
normalized Fluc/Rluc (A.U.)
G
actin(inv)
100 nt
P4-actin(inv)
P4-actin(inv)
actin(inv)
P4(Mx)-actin(inv)
normalized Fluc/Rluc (A.U.)

Figure S1. The short stem-loop P4 in the *Hoxa9* 5' UTR is sufficient for recruitment of the ribosome and acts as a translational enhancer. Related to Figure 1, 2.

(**A**) RT-PCR analysis of the *Hoxa9* IRES-like element using cDNA from stage E11.5 FVB embryo mRNA and primers specific for the most 5' region of the IRES in the 5' UTR and for the CDS region (see **Table S3**). Sequencing of the single PCR product confirms the *a9* IRES sequence as annotated.

(**B**) Schematic representation of the mouse *Hoxa9* secondary structure model of the 180 nucleotides (nt) long *Hoxa9* IRES-like element RNA (termed *a9* $IRES_{180}$) (Xue et al., 2015) containing four pairing elements P1-P4 and a putative pseudoknot (PK), and P4 (red) highlighted. Numbers refer to nucleotide positions within the *Hoxa9* 5' UTR.

(**C**) Secondary structure model of the P4 stem-loop and substitution mutations mapped onto the P4 structure. Numbers refer to nucleotide positions within the *Hoxa9* 5' UTR. P4 stem mutations were introduced either in one strand to disrupt P4, or as compensatory mutations in both strands to restore P4. P4 mutants active in mediating IRES-like activity in the context of the fusion to the native spacer (P4-native) or the *actin*(inv) 5' UTR sequence (P4-*actin*(inv)) (normalized Fluc/Rluc < 0.5 A.U.) are labeled "+", moderately active mutants (Fluc/Rluc < 0.5, > 1.0 A.U.) are labeled "+/–", and inactive mutants (Fluc/Rluc > 0.5 A.U.) are labeled "–". Sequence critical for IRES activity is highlighted in yellow. Adapted from **Figure 1C**.

(**D**) The same substitution mutations as in (C) were mapped onto the linear P4 sequence, together with the IRES-like activity (+, +/–, –) of the corresponding reporter mRNAs in context of the native spacer or the *actin*(inv) sequence.

(**E**) Secondary structure models of WT P4 and P4(M5) were based on SHAPE structure probing (Xue et al., 2015) and prediction of structural changes induced by the 4 nt M5 mutation (green) using Vienna RNAfold (http://rna.tbi.univie.ac.at) and visualized using VARNA (http://varna.lri.fr) with default settings.

(**F**) The effect of the P4 substitution mutations in the P4-native context, as displayed in the reporter schematic, on IRES-like activity was measured by transiently transfecting mouse C3H/10T1/2 cells with the corresponding bicistronic reporter genes containing no insert in the intergenic region (pRF, empty vector), *a9* IRES FL, P4, native or a fusion of P4 mutants M1-M11 with the native spacer downstream of P4 (labelled M1-M11). Cells were harvested after 24 hours for protein lysates and subjected to the luciferase activity assay. Luciferase activity analysis was carried out as in **Figure 1E**. Relative luciferase activity was expressed as a Fluc(IRES)/Rluc(cap-initiation) ratio and expressed as average IRES activity ± SEM, n = 3-8; ns, not significant. FL, full-length.

(**G**) Bicistronic reporter assay as described in (E), and as displayed in the reporter schematic, using P4-*actin*(inv) fusion constructs testing P4 mutants M1-M11. Relative luciferase activity was expressed as average IRES activity ± SEM, n = 3-5.

(**H**) Schematic representation of the mouse *Hoxa9* IRES-like element, including P4 and the "native" spacer of 130 nt between P4 and the AUG. Spacer length requirements for P4 IRES-like activity were tested by insertion of spacers of different lengths, the native spacer, the inverse *actin* 5' UTR sequence (*actin*(inv)), or the HBB 5' UTR (*HBB*), downstream of P4 in a bicistronic reporter mRNA. Rluc, renilla luciferase; Fluc, firefly luciferase. FL, full-length. Partially reproduced from **Figure 1B, D**.

(**I**) Luciferase activity analysis in C3H/10T1/2 cells was carried out as in **Figure 1E**. Relative luciferase activity is expressed as a Fluc(IRES)/Rluc(cap-initiation) ratio normalized to respective Fluc/Rluc mRNA levels and given as average IRES activity ± SEM, n = 5-9; ns, not significant.

(**J**) Cap inhibition by eIF4E-binding drug 4EGI-1 uncouples cap-dependent translation from P4-mediated translation enhancer function. Schematic of m^7G-capped Fluc and Nluc reporter mRNAs and inhibitor function. See also **Figure 2E**.
(**K**) Luciferase activity analysis in C3H/10T1/2 cells was carried out as in **Figure 2F**. Luciferase activity values that correspond to the data in **Figure 2F** are expressed as normalized Fluc or Nluc luminescence given as average luminescence ± SEM, n = 4; ns, not significant; *actin*(inv), untreated was set to 1.

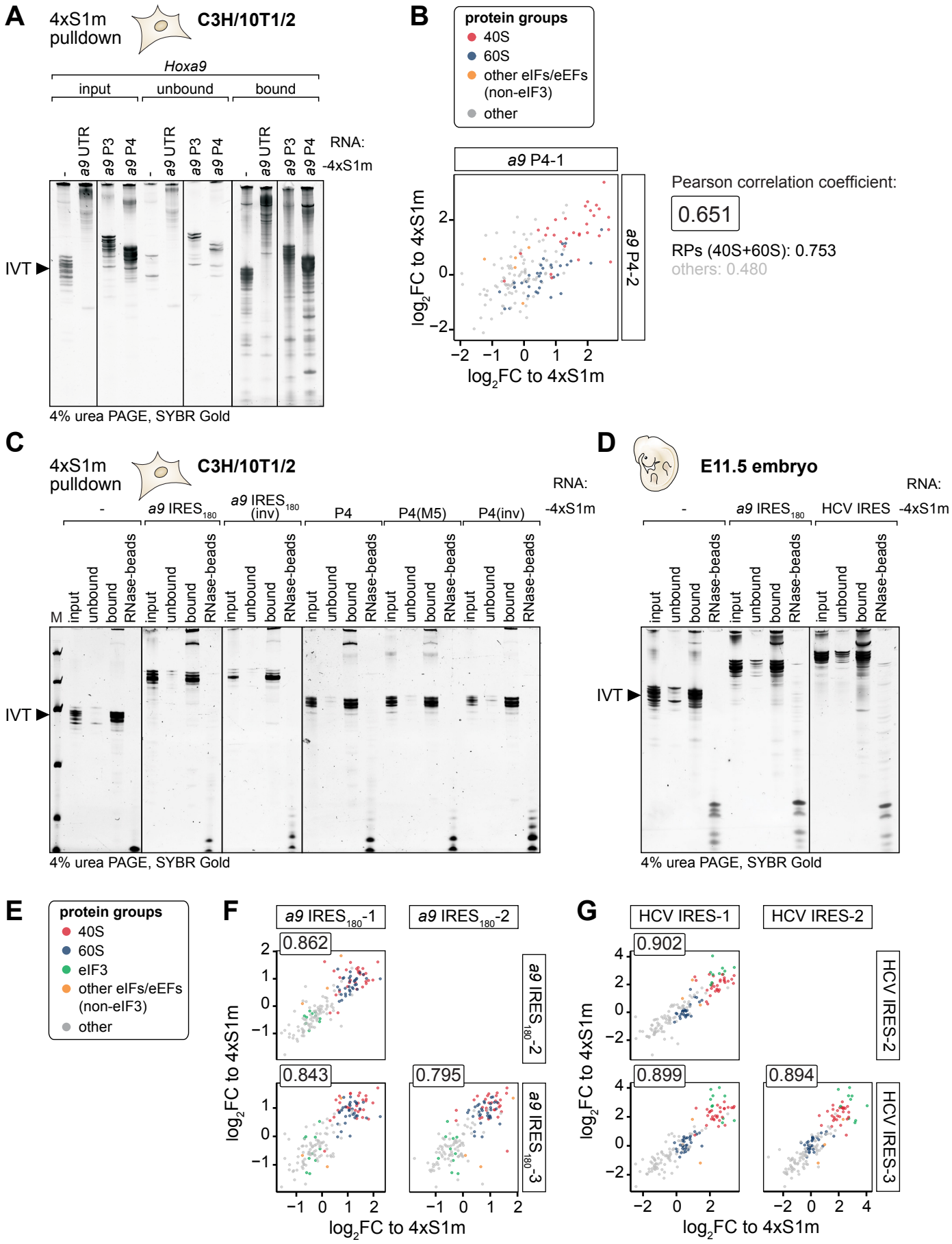
A
4xS1m pulldown
C3H/10T1/2
Hoxa9
input
unbound
bound
- a9 UTR a9 P3 a9 P4 - a9 UTR a9 P3 a9 P4 - a9 UTR a9 P3 a9 P4
RNA: -4xS1m
IVT
4% urea PAGE, SYBR Gold
B
protein groups
40S
60S
other eIFs/eEFs (non-eIF3)
other
a9 P4-1
a9 P4-2
log2FC to 4xS1m
Pearson correlation coefficient:
0.651
RPs (40S+60S): 0.753
others: 0.480
C
4xS1m pulldown
C3H/10T1/2
RNA: -4xS1m
- a9 IRES180 a9 IRES180 (inv) P4 P4(M5) P4(inv)
M input unbound bound RNase-beads
IVT
4% urea PAGE, SYBR Gold
D
E11.5 embryo
RNA: -4xS1m
- a9 IRES180 HCV IRES
input unbound bound RNase-beads
IVT
4% urea PAGE, SYBR Gold
E
protein groups
40S
60S
eIF3
other eIFs/eEFs (non-eIF3)
other
F
a9 IRES180-1
a9 IRES180-2
a9 IRES180-3
0.862
0.843
0.795
log2FC to 4xS1m
G
HCV IRES-1
HCV IRES-2
HCV IRES-3
0.902
0.899
0.894
log2FC to 4xS1m

Figure S2. The *Hoxa9* IRES-like element binds to the 40S ribosomal subunit via P4 and to the 80S ribosome. Related to Figure 3.

(**A**) *In vitro* transcribed RNAs fused to 4xS1m aptamers were coupled to SA-sepharose beads for 4xS1m pulldown using untransfected C3H/10T1/2 cells as input. Coupled beads were incubated with cell extracts, washed and eluted using RNase A to release RNA-bound proteins. Input and unbound samples were taken before and after incubation of RNAs with beads. After incubation with extracts and wash steps, and before RNase A-elution, a "bound" fraction of the beads for each sample was saved. To monitor coupling efficiency, 10% of the input, unbound, and bound RNA fraction of each sample was resolved by 4% denaturing polyacrylamide/TBE/urea PAGE and visualized by SYBR Gold. Representative of n = 2 is shown. Corresponds to **Figure 3B**.

(**B**) Correlation plot of TMT-MS/MS data from C3H/10T1/2 cells. Comparison of the pair of individual TMT-MS/MS samples (two replicate samples per RNA *a9* P4 bait). Scatter plot of $\log_2$ fold change (FC) relative to the aptamer control (4xS1m) for *a9* P4, colored by protein group (see legend) and Pearson correlation coefficient for the whole data set and for RPs and other proteins. See also **Figure 3D**, **Table S4**.

(**C**) 4xS1m pulldown with *a9* $IRES_{180}$- and P4-4xS1m as well as control constructs was performed as described in panel (A). Analysis of RNA fractions of beads after RNase A-treatment (RNase-beads) confirm complete digestion of coupled RNAs on beads. Representative of n = 3 is shown. Corresponds to **Figure 3E**.

(**D**) The same 4xS1m pulldown experiment was carried out as described in panel (A), except that stage E11.5 FVB mouse embryos were used to generate cellular extracts. Analysis of RNA fractions of beads after RNase A-treatment (RNase-beads) confirm complete digestion of coupled RNAs on beads. Representative of n = 3 is shown. Corresponds to **Figure 3F**.

(**E**) Protein groups for analysis of reproducibility of TMT-MS/MS data in (F) and (G).

(**F**) Correlation matrix of TMT-MS/MS data from stage E11.5 FVB mouse embryos. A matrix comparing every possible pair of individual TMT-MS/MS samples (three replicate samples per *a9* $IRES_{180}$ RNA bait). Scatter plots of $\log_2$ fold change (FC) relative to the aptamer control (4xS1m) for *a9* $IRES_{180}$, colored by protein group (E) and Pearson correlation coefficient for the whole data set is shown. See also **Figure 3G**, **Table S5**.

(**G**) Correlation matrix of TMT-MS/MS data from stage E11.5 FVB mouse embryos as in (F) for three replicate samples per HCV IRES RNA bait. See also **Figure 3G**, **Table S5**.

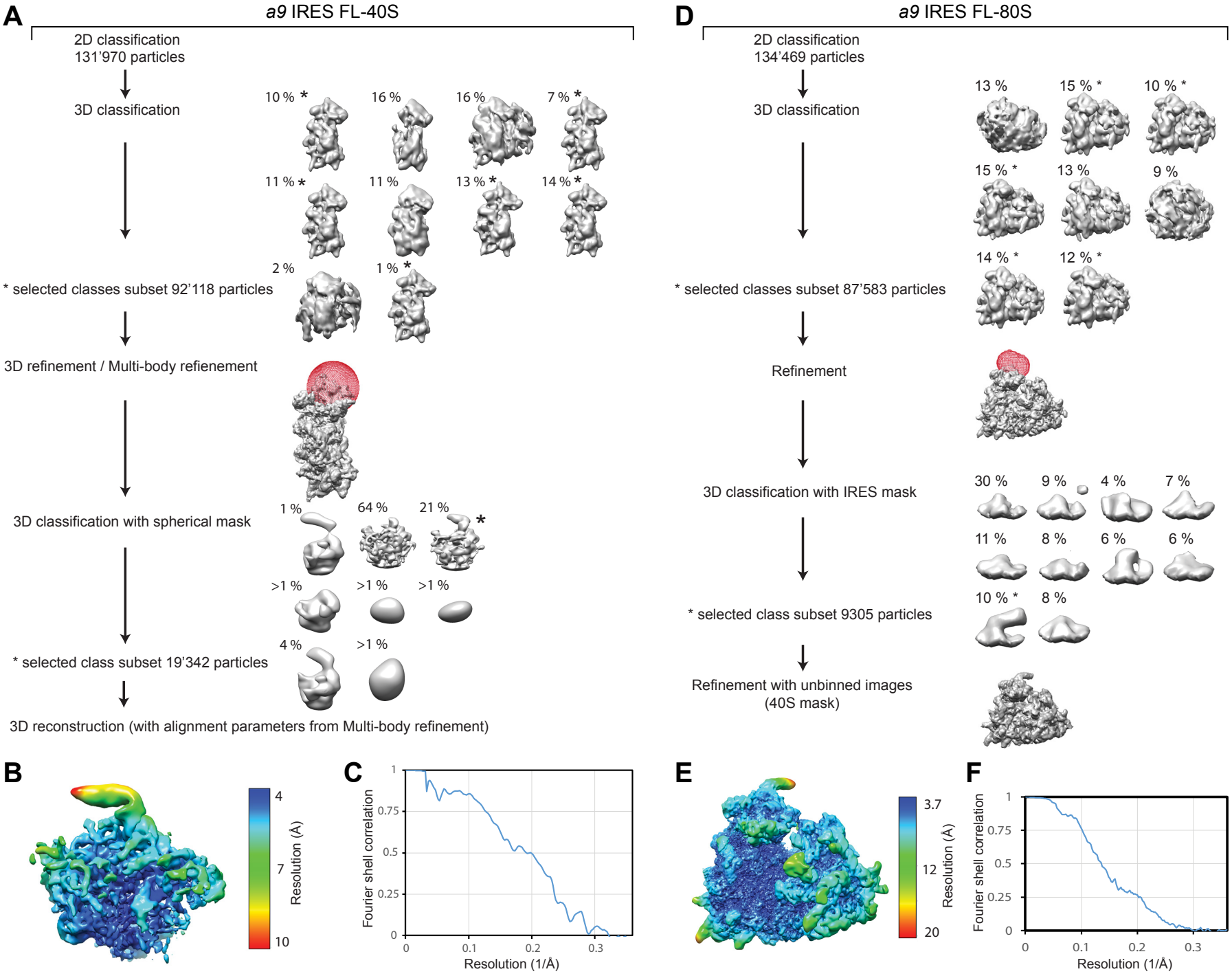
A
a9 IRES FL-40S
2D classification
131'970 particles
3D classification
10 % *
16 %
16 %
7 % *
11 % *
11 %
13 % *
14 % *
2 %
1 % *
* selected classes subset 92'118 particles
3D refinement / Multi-body refienement
3D classification with spherical mask
1 %
64 %
21 % *
>1 %
>1 %
>1 %
4 %
>1 %
* selected class subset 19'342 particles
3D reconstruction (with alignment parameters from Multi-body refinement)
D
a9 IRES FL-80S
2D classification
134'469 particles
3D classification
13 %
15 % *
10 % *
15 % *
13 %
9 %
14 % *
12 % *
* selected classes subset 87'583 particles
Refinement
3D classification with IRES mask
30 %
9 %
4 %
7 %
11 %
8 %
6 %
6 %
10 % *
8 %
* selected class subset 9305 particles
Refinement with unbinned images
(40S mask)
B
4
7
10
Resolution (Å)
C
Fourier shell correlation
1
0.75
0.5
0.25
0
0
0.1
0.2
0.3
Resolution (1/Å)
E
3.7
12
20
Resolution (Å)
F
Fourier shell correlation
1
0.75
0.5
0.25
0
0
0.1
0.2
0.3
Resolution (1/Å)

Figure S3. Cryo-EM classification and refinement of the *a9* IRES FL in complex with 40S and 80S. Related to Figure 4.

(**A**) Cryo-EM image classification and refinement for the *a9* IRES FL-40S subunit complex. Schematic representation of the 2D and 3D classification steps used to obtain the *a9* IRES FL-40S subunit complex. Corresponds to **Figure 4A**.

(**B**) Density of the 40S subunit head in complex with the *a9* IRES FL filtered to local resolution and color coded by local resolution value.

(**C**) Fourier shell resolution curve of the final 3D reconstruction of the *a9* IRES FL-40S subunit complex head density indicating an overall resolution of 3.9 Å.

(**D**) Cryo-EM image classification and refinement for the *a9* IRES FL-80S subunit complex. Schematic representation of the 2D and 3D classification steps used to obtain the *a9* IRES FL-80S subunit complex. Corresponds to **Figure 4B**.

(**E**) Density of the 80S ribosome in complex with the *a9* IRES FL filtered to local resolution and color coded by local resolution value.

(**F**) Fourier shell resolution curve of the final 3D reconstruction of the *a9* IRES FL-80S subunit complex head density indicating an overall resolution of 4.4 Å.

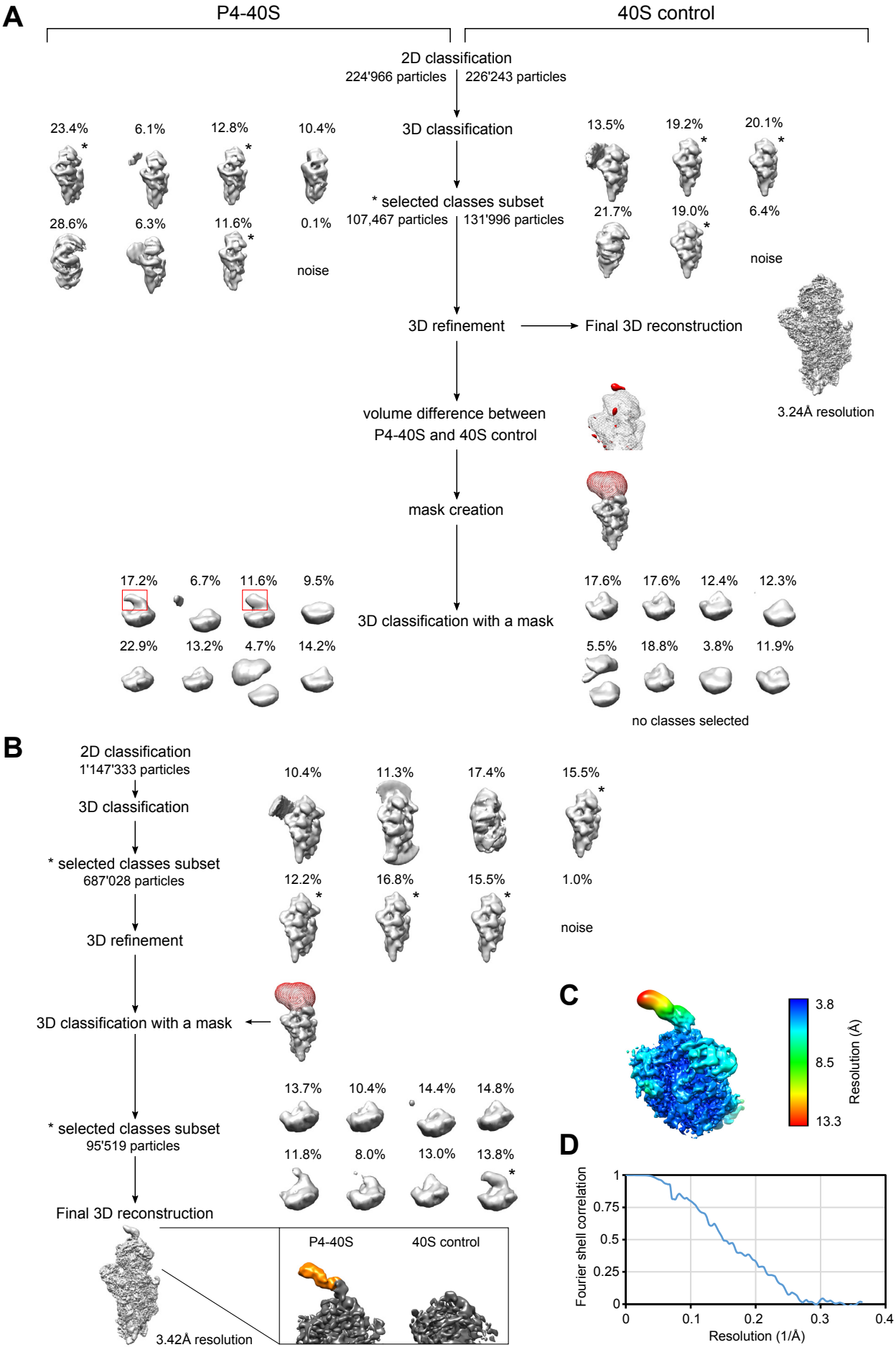
A
P4-40S
40S control
2D classification
224'966 particles
226'243 particles
3D classification
23.4%
6.1%
12.8%
10.4%
13.5%
19.2%
20.1%
* selected classes subset
107,467 particles
131'996 particles
28.6%
6.3%
11.6%
0.1%
noise
21.7%
19.0%
6.4%
noise
3D refinement
Final 3D reconstruction
3.24Å resolution
volume difference between
P4-40S and 40S control
mask creation
3D classification with a mask
17.2%
6.7%
11.6%
9.5%
22.9%
13.2%
4.7%
14.2%
17.6%
17.6%
12.4%
12.3%
5.5%
18.8%
3.8%
11.9%
no classes selected
B
2D classification
1'147'333 particles
3D classification
* selected classes subset
687'028 particles
3D refinement
3D classification with a mask
* selected classes subset
95'519 particles
Final 3D reconstruction
10.4%
11.3%
17.4%
15.5%
12.2%
16.8%
15.5%
1.0%
noise
13.7%
10.4%
14.4%
14.8%
11.8%
8.0%
13.0%
13.8%
P4-40S
40S control
3.42Å resolution
C
3.8
8.5
13.3
Resolution (Å)
D
Fourier shell correlation
1
0.75
0.5
0.25
0
0
0.1
0.2
0.3
0.4
Resolution (1/Å)

Figure S4. Cryo-EM classification and refinement of the Hox*a9* IRES-P4 in complex with 40S and 40S control. Related to Figure 4.

(**A**) Processing scheme of the of P4-40S and 40S control cryo-EM datasets with similar number of particles upon 2D classification. Red squares indicate P4 density protrusions in P4-40S dataset which are not visible in the 40S control dataset.

(**B**) Processing scheme of the of whole P4-40S cryo-EM dataset. Further processing steps included multibody refinement, followed by a 3D classification on a P4 sub-region using a tighter mask. Surface representation of a final reconstruction, filtered to 7Å resolution, clearly shows RNA features of the P4 on the 40S (colored orange) which are absent in control reconstruction of a 40S ribosomal subunit from human cells. Corresponds to **Figure 4C-E**.

(**C**) Density of the 40S ribosomal subunit in complex with *Hoxa9* IRES P4 element filtered to local resolution and color coded by local resolution value.

(**D**) Fourier shell resolution curve of the *Hoxa9* IRES P4-40S head reconstruction indicating an overall resolution of 4.1 Å.

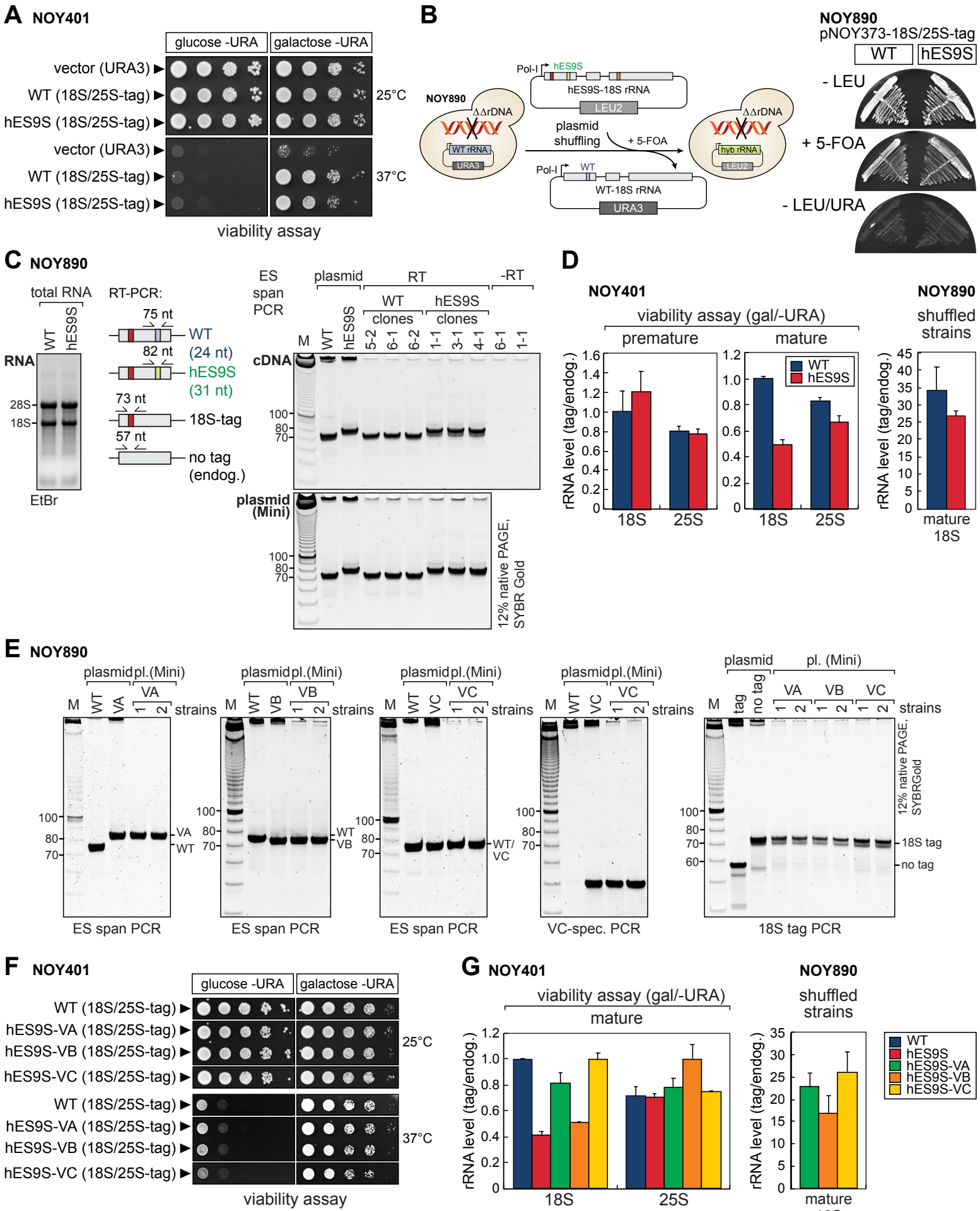
A NOY401
glucose -URA
galactose -URA
vector (URA3)
WT (18S/25S-tag)
hES9S (18S/25S-tag)
25°C
37°C
viability assay
B
NOY890
pNOY373-18S/25S-tag
WT
hES9S
- LEU
+ 5-FOA
- LEU/URA
Pol-I
hES9S
hES9S-18S rRNA
LEU2
ΔΔrDNA
WT rRNA
URA3
plasmid shuffling
+ 5-FOA
hyb rRNA
WT
WT-18S rRNA
C NOY890
total RNA
RT-PCR:
RNA
28S
18S
EtBr
75 nt
WT (24 nt)
82 nt
hES9S (31 nt)
73 nt
18S-tag
57 nt
no tag (endog.)
ES span PCR
plasmid
RT
-RT
WT clones
hES9S clones
cDNA
plasmid (Mini)
12% native PAGE, SYBR Gold
D
NOY401
NOY890
viability assay (gal/-URA)
premature
mature
shuffled strains
rRNA level (tag/endog.)
18S
25S
mature 18S
E NOY890
plasmid
pl.(Mini)
VA
VB
VC
strains
ES span PCR
VC-spec. PCR
18S tag PCR
tag
no tag
12% native PAGE, SYBRGold
18S tag
no tag
F NOY401
glucose -URA
galactose -URA
WT (18S/25S-tag)
hES9S-VA (18S/25S-tag)
hES9S-VB (18S/25S-tag)
hES9S-VC (18S/25S-tag)
25°C
37°C
viability assay
G NOY401
NOY890
viability assay (gal/-URA)
mature
shuffled strains
rRNA level (tag/endog.)
18S
25S
mature 18S
WT
hES9S
hES9S-VA
hES9S-VB
hES9S-VC

Figure S5. Plasmid shuffling and yeast strain characterization. Related to Figure 5.

(**A**) Viability of engineered yeast strains containing hES9S in their 18S rRNA was assayed in the NOY401 strain harboring a temperature-sensitive RNA polymerase I allele (pol I). The plasmid encoding hES9S-rDNA contains a uracil (*URA3*) marker. hES9S-rDNA expression is controlled by a GAL-promoter. Induced expression on galactose plates at restrictive temperature (37°C) supports growth of cells that can survive based on plasmid-provided rRNA. A *URA3* plasmid and a yeast WT rDNA plasmid serve as negative and positive controls, respectively. All rDNA plasmids contain 18S and 25S rRNA tags. Representative of n = 3 is shown.

(**B**) (Right) Schematic of the plasmid shuffling approach to generate yeast strains (NOY890) that contain a homozygous knock-out of the rDNA locus and generate ribosomes exclusively from plasmids. Adapted from **Figure 5E**. (Left) Plasmid exchange from *URA3* (WT) to *LEU2* (tagged WT or hES9S)-plasmids in isolates is tested by growth on drop-out plates. Successful plasmid shuffling results in growth of isolates on SD-*LEU2*, and SD+5-FOA but not on SD-*LEU/URA*. Exemplary clones of WT and hES9S-containing rRNA strains were streaked out on respective plates and their growth was documented.

(**C**) RT-PCR analysis using ES9S-specific primers that span ES9S allow analysis of expression of WT or hES9S 18S rRNA due to a PCR product of 7 nt difference in length between WT and hES9S (ES span PCR). Similarly, the presence of the 18S tag can be distinguished from WT rRNA (18S tag PCR). Total RNA for cDNA synthesis or plasmid DNA was extracted from clones and used for RT-PCR. Plasmid-derived PCR products serve as controls. PCR products were resolved by 12% native PAGE and stained with SYBR Gold. Three clones (NOY890) used in this study are presented. 2 μg total RNA of WT and hES9S strains was resolved on a 1.2% denaturing agarose gel to assess rRNA processing. A 10 bp DNA ladder (Invitrogen) was loaded as reference.

(**D**) Yeast cells (NOY401) for the viability assay (A) were harvested at OD_{600} = 1.0 after overnight induction in galactose-media at RT and subjected to total RNA extraction (n = 3, SD). For the viability assay, RT-qPCR analysis using specific primers for rRNA tags and endogenous rRNA assesses the degree of processing from premature to mature rRNA for 18S and 25S. For yeast strain characterization after plasmid shuffling and isolation of clones, this tag/endogenous rRNA ratio assesses the substitution rate of WT with tagged-WT or tagged-hES9S ribosomes present in isolated strains. For NOY890 strains, for 33 and 26 tagged WT and hES9S ribosomes, respectively, one endogenous plasmid-derived WT 40S ribosome is left in the cell.

(**E**) Same RT-PCR analysis as in (C) was performed using ES9S-specific primers that span ES9S (ES span PCR) and primers for the 18S tag (18S tag PCR) for the characterization of the hES9S-variant strains VA-VC. Plasmid DNA was extracted from clones and used for RT-PCR. Plasmid-derived PCR products serve as controls. Two clones (NOY890) used in this study are presented. A 10 bp DNA ladder (Invitrogen) was loaded as reference.

(**F**) Viability of engineered yeast strains containing hES9S-VA, -VB, or -VC in their 18S rRNA was assayed in the NOY401 strain harboring a temperature-sensitive RNA pol I as in (A). A yeast WT rDNA plasmid serve as positive control. All rDNA plasmids contain 18S and 25S rRNA tags. Representative of n = 3 is shown.

(**G**) Yeast cells (NOY401) for the viability assay (F) were harvested at OD_{600} = 1.0 after overnight induction in galactose-media at RT and subjected to total RNA extraction (n = 3, SD) as in (D). For the viability assay, RT-qPCR analysis using specific primers for rRNA tags and endogenous rRNA assesses the degree of processing to mature rRNA for 18S and 25S. For NOY890 strains after plasmid shuffling, for 22, 17 and 24 tagged hES9S-VA, -VB and -VC ribosomes, respectively, one endogenous plasmid-derived WT 40S ribosome is left in the cell.

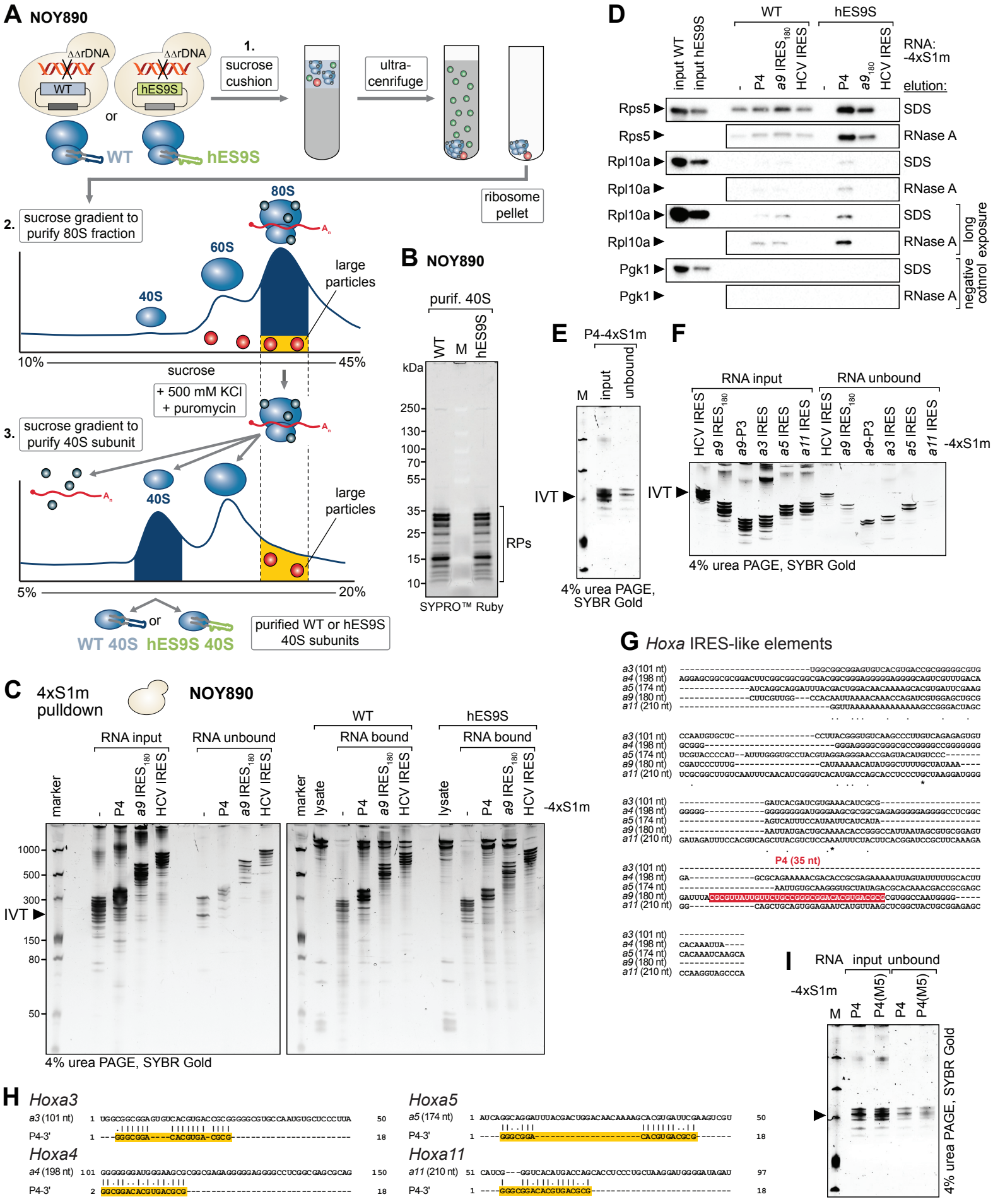
A NOY890
ΔΔrDNA
WT
hES9S
or
1. sucrose cushion
ultra-cenrifuge
ribosome pellet
2. sucrose gradient to purify 80S fraction
80S
60S
40S
large particles
10%
sucrose
45%
+ 500 mM KCl + puromycin
3. sucrose gradient to purify 40S subunit
5%
20%
purified WT or hES9S 40S subunits
WT 40S hES9S 40S
B NOY890
purif. 40S
WT M hES9S
kDa
250
130
100
70
55
35
25
15
10
RPs
SYPRO™ Ruby
C 4xS1m pulldown
NOY890
RNA input
RNA unbound
RNA bound
WT
hES9S
marker
lysate
-
P4
a9 IRES180
HCV IRES
-4xS1m
1000
500
300
IVT
150
80
50
4% urea PAGE, SYBR Gold
D
input WT
input hES9S
WT
hES9S
RNA: -4xS1m
elution:
Rps5
Rpl10a
Pgk1
SDS
RNase A
long exposure
negative cotnrol
E P4-4xS1m
M
input
unbound
IVT
4% urea PAGE, SYBR Gold
F
RNA input
RNA unbound
HCV IRES
a9 IRES180
a9-P3
a3 IRES
a5 IRES
a11 IRES
-4xS1m
IVT
4% urea PAGE, SYBR Gold
G Hoxa IRES-like elements
a3 (101 nt)
a4 (198 nt)
a5 (174 nt)
a9 (180 nt)
a11 (210 nt)
P4 (35 nt)
H Hoxa3
a3 (101 nt) 1 UGGCGGCGGAGUGUCACGUGACCGCGGGGGCGUGCCAAUGUGCUCCCUUA 50
P4-3' 1 GGGCGGA----CACGUGA-CGCG 18
Hoxa4
a4 (198 nt) 101 GGGGGGGAUGGGAAGCGCGGCGAGAGGGGGAGGGCCUCGGCGAGCGCAG 150
P4-3' 2 GGCGGACACGUGACGCG 18
Hoxa5
a5 (174 nt) 1 AUCAGGCAGGAUUUACGACUGGACAACAAAAGCACGUGAUUCGAAGUCGU 50
P4-3' 1 GGGCGGA----------CACGUGACGCG 18
Hoxa11
a11 (210 nt) 51 CAUCG---GGUCACAUGACCAGCACCUCCCUGCUAAGGAUGGGGAUAGAU 97
P4-3' 1 GGGCGGACACGUGACGCG 18
I
RNA input unbound
-4xS1m
M P4 P4(M5) P4 P4(M5)
4% urea PAGE, SYBR Gold

Figure S6. Purification of WT and hES9S 40S ribosomal subunits from yeast and 4xS1m pulldown of yeast WT and hES9S ribosomes with IRES-like elements. Related to Figure 4, 6, 7.

(**A**) Purification scheme for 40S ribosomal subunit isolation from WT and hES9S yeast strains (NOY890) by sequential sucrose cushion (1.) and gradient centrifugation (2. and 3.). First, ribosomes from yeast cell lysates derived from either WT or hES9S humanized yeast strains were pelleted by ultracentrifugation on a sucrose cushion (1.), the ribosome pellet is loaded onto a 10-45% sucrose gradient (2.) and the 80S ribosomal fractions are treated with high salt and puromycin to split them into 40S and 60S. A second 5-20% gradient fractionation (3.) of individual subunits separated 40S subunits from free mRNA and proteins, 60S, and large particles.

(**B**) SDS-PAGE analysis and SYPRO-Ruby staining of purified 40S ribosomal subunits from WT and hES9S yeast strains indicates similar purity and ribosomal protein (RP) enrichment in both 40S fractions (10-35 kDa range). See also corresponding WB analysis in **Figure 5F**.

(**C**) A similar 4xS1m pulldown experiment was carried out as described in **Figure S2A**, except that tagged-WT or tagged-hES9S ribosome expressing yeast strains were used to generate cellular extracts. Input and unbound RNA samples were taken before and after incubation of RNAs with beads. To monitor coupling efficiency, 10% of the input and unbound RNA fraction of each sample was resolved by 4% denaturing polyacrylamide/TBE/urea PAGE and visualized by SYBR Gold staining. Representative of n = 3 is shown. Corresponds to **Figure 6B**. Low Range ssRNA Ladder (NEB) was loaded for reference.

(**D**) The extended data for the WB in **Figure 6B** and the corresponding RNA gel in (C) demonstrates the improved specificity of the RNase A elution over elution of the beads by SDS-containing sample buffer (SDS) in the 4xS1m pulldown experiment. PGK1 serves as a negative control. Representative of n = 3 is shown.

(**E**) A similar 4xS1m pulldown experiment as described in (C) was performed with the focus on the hES9S-variant VA-VC comparison and 4% urea PAGE analysis of RNA input and unbound samples. Corresponds to **Figure 6D**.

(**F**) A similar 4xS1m pulldown experiment with the focus on the *Hoxa* IRES-like comparison and 4% urea PAGE analysis of RNA input and unbound samples was performed as described in (C). Corresponds to **Figure 6E**.

(**G**) Alignment of full-length *Hoxa* IRES-like element sequences. IRES lengths are given in nts. The P4 in *a9* is indicated in red.

(**H**) Alignments of the P4-3' motif (indicated in yellow) to individual full-length *Hoxa* IRES sequences.

(**I**) A similar 4xS1m pulldown experiment with the focus on the comparison of P4 to P4(M5) and 4% urea PAGE analysis of RNA input and unbound samples was performed as described in (C). Representative of n = 4 is shown. Corresponds to **Figure 7B**.

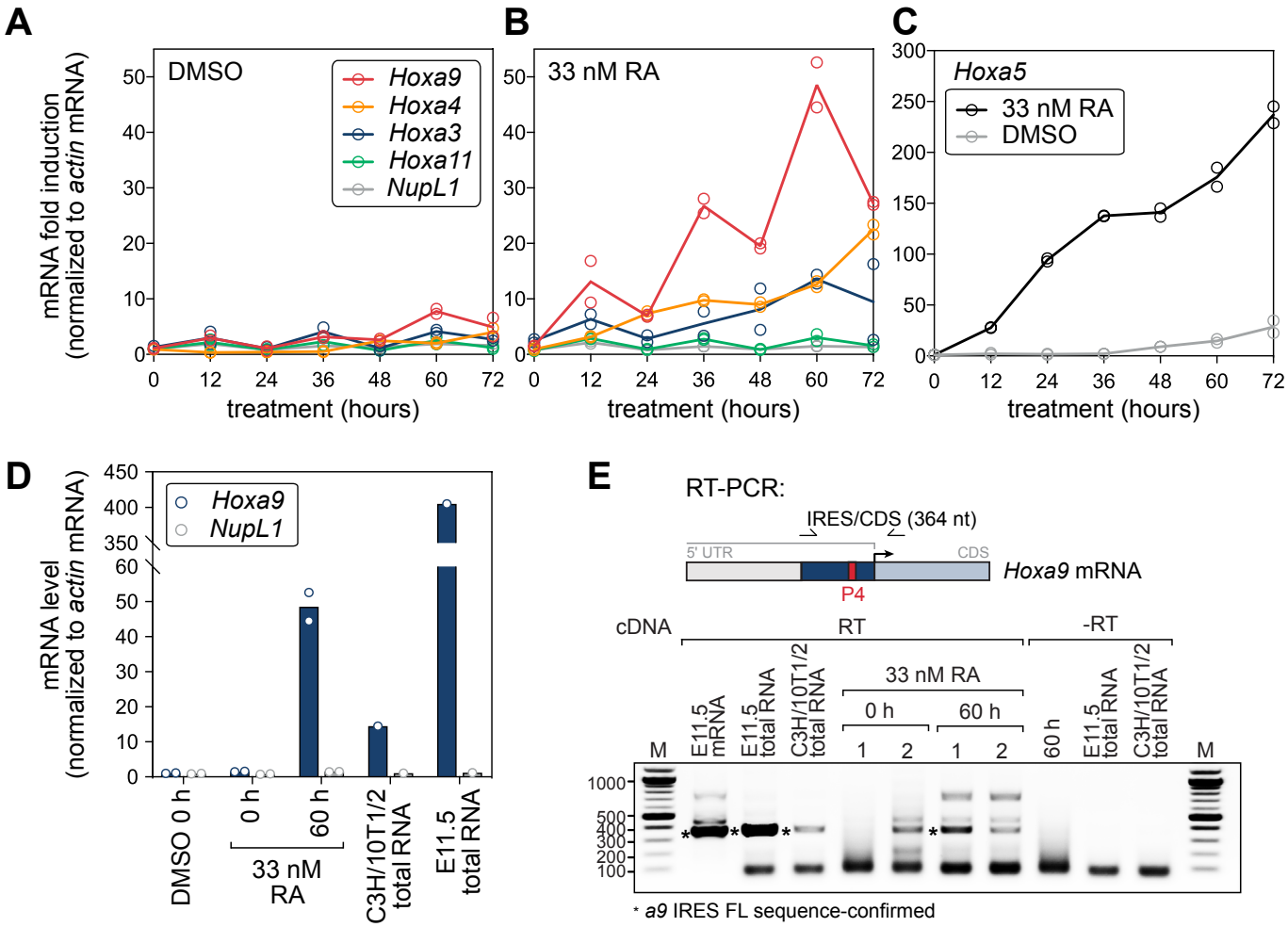
A
DMSO
Hoxa9
Hoxa4
Hoxa3
Hoxa11
NupL1
mRNA fold induction (normalized to actin mRNA)
treatment (hours)
B
33 nM RA
treatment (hours)
C
Hoxa5
33 nM RA
DMSO
treatment (hours)
D
Hoxa9
NupL1
mRNA level (normalized to actin mRNA)
DMSO 0 h
0 h
60 h
33 nM RA
C3H/10T1/2 total RNA
E11.5 total RNA
E
RT-PCR:
IRES/CDS (364 nt)
5' UTR
CDS
Hoxa9 mRNA
P4
cDNA
RT
-RT
33 nM RA
0 h
60 h
E11.5 mRNA
E11.5 total RNA
C3H/10T1/2 total RNA
M
1
2
* a9 IRES FL sequence-confirmed
F
WT
edited
P4
P4(M5)
sgRNA-1
Cas9
PAM
a9 IRES180
cut site
M5-specific primer
TTACGCGTTATTGTTCTGCCGGGCGGACACGTGACGCGCGTGGCCAATGGGGGCGCGGGCGCCG
TTACGCGTTATTGTTCTGCCGGGCGGACACTATTCGCGCGTGGCCAATGGGGGCGCGGGCGCCG
ssDNA template (30 nt or 60 nt overhang)
M5-specific PCR:
a9 IRES FL
M5-specific (219 nt)
CDS
Hoxa9
P4
a9 IRES180
IRES/CDS PCR:
IRES/CDS (364 nt)
gDNA
clones
E14 (WT)
A7
D6
clone D6 (homozygous)
GGGCGGACACTATTCGCGCGTGGC

Figure S7. Induction of HoxA genes in mESCs by retinoic acid treatment and CRISPR/Cas9-editing of M5 into the HoxA9 gene. Related to Figure 7.

(**A**) For Hox gene induction, mES cells were treated with DMSO alone as a negative control (A) or with 33 nM retinoic acid (RA) in DMSO (**B**, **C**). An RA concentration of 33 nM closely mimicked physiological oscillation of Hox gene induction (De Kumar et al., 2015). Fresh media with RA was provided every 12 hours and cells were harvested in 12 hour-intervals for 0-72 hours for time course experiments, and subjected to RNA extraction and RT-qPCR. mRNA fold induction is expressed as mRNA levels of *Hoxa* or *NupL1* as a control normalized to respective *actin* mRNA levels. Sample DMSO, 0 hours was set to 1 for each mRNA, n = 2.

(**D**) Comparison of expression levels of *Hoxa9* and *NupL1* mRNAs relative to *actin* mRNA in RA or DMSO-treated mES cells, C3H/10T1/2 cells, and stage E11.5 FVB mouse embryos using DNase-treated total RNA as input. Sample DMSO, 0 hours was set to 1, n = 1-2.

(**E**) RT-PCR analysis of the *Hoxa9* IRES in the 5' UTR of *Hoxa9* mRNA as induced by RA-treatment. DNase-treated total RNA or mRNA as in (D) was used for RT using random hexamers and primers flanking the IRES 5' and priming off the *a9* CDS 3', respectively. This resulted in a PCR product of 364 nts as indicated in the schematic. Bands indicated with asterisks were gel-extracted and sequenced and confirm that the *a9* mRNA induced upon 60 hours of RA-treatment contains the IRES as annotated, identical to in mouse embryos and C3H/10T1/2 cells. For 0 and 60 hours of RA, two biological replicates were tested (1, 2). Primers are given in **Table S3**.

(**F**) Schematic of the targeted CRISPR/Cas9 genome editing of P4(M5) into the *Hoxa9* locus in mESCs using guide RNAs that cut very close to the M5 site and single-stranded DNA donor templates (ssDNA) that have either 30 or 60 nt overhangs 5' and 3' of the 4-nt M5 mutation (TATT). We used three different guide RNAs, of which sgRNA-1 was the most efficient (cut site depicted). Two genotyping strategies were designed to characterize successful editing in isolated clones after genomic DNA (gDNA) extraction. The first PCR amplifies only M5-edited regions (M5-specific PCR, 219 nt, KL606/597) whereby the forward primer primes off the 4-nt M5 mutation and only yields a product if editing has occurred. The second PCR is a M5-spanning PCR (IRES/CDS PCR, 364 nt, KL596/597) that contains the edit site in the center and should always yield a product and was used for sequence confirmation by Sanger sequencing using both outer primers. This PCR strategy identified two positive clones in comparison to unedited WT cells (E14) of which D6 is a homozygous clone generated with sgRNA-1 and the 60 nt-overhang ssDNA. The chromatogram after Sanger sequencing is shown for clone D6 indicating scarless, homozygous editing of M5 (TATT) into the *Hoxa9* locus.

SUPPLEMENTAL TABLES

Table S1: Plasmids used in this study. Related to STAR Methods.

All plasmids used for *in vitro* transcription and mammalian transient transfection or yeast transformation are listed in the table.

Table S1. List of plasmids

Plasmid	Notes	Reference
***In vitro* transcription constructs**		
pSP73	SP6 promoter, kindly provided by G. Stoecklin	Promega
pSP73-4xS1m	p2880, kindly provided by G. Stoecklin	(Leppek and Stoecklin, 2014)
pSP73-4xS1m(MCS)		This study
pSP73-a9(5' UTR)-4xS1m(MCS)		This study
pSP73-a9(IRES)-4xS1m(MCS)		This study
pSP73-a9(P3)-4xS1m(MCS)		This study
pSP73-a9(P4)-4xS1m(MCS)		This study
pSP73-a3(IRES)-4xS1m(MCS)		This study
pSP73-a5(IRES)-4xS1m(MCS)		This study
pSP73-a9(IRES)$_{180}$-4xS1m(MCS)		This study
pSP73-a11(IRES)-4xS1m(MCS)		This study
pSP73-HCV IRES-4xS1m(MCS)		This study
pSP73-a9(P4(M5))-4xS1m(MCS)		This study
Mammalian cells		
Expression constructs		
pRF	SV40 promoter, kindly provided by D. Ruggero	
pRF-HCV IRES	kindly provided by D. Ruggero	
pRF-EMCV IRES	kindly provided by D. Ruggero	
pRL	SV40 promoter, Renilla luciferase	Promega
pGL3	SV40 promoter, Firefly luciferase	Promega
pGL3-HBB		(Xue et al., 2015)
pRF-a9 5' UTR		(Xue et al., 2015)
pRF-a9-IRES FL		(Xue et al., 2015)
pFLB (pcDNA3-Fluc-β-globin)	p2524, CMV promoter	(Ozgur et al., 2010)
pRF-actin	mouse β-*actin* 5' UTR	This study
pRF-a9-IRES$_{180}$		This study
pRF-actin-a9-IRES$_{180}$		This study
pRF-a9-IRES$_{180}$-actin		This study
pRF-a9-IRES$_{180}$(M5)-actin		This study
pRF-a9-P3		This study
pRF-a9-P3-actin		This study
pRF-a9-P4		This study
pRF-a9-P4-actin		This study

pRF-a9-P4(M5)-actin		This study
pRF-a9-P4-native		This study
pRF-a9-P4(M1-M11)-native	P4 mutagenesis series M1-M11	This study
pRF-a9-native		This study
pRF-actin(inv)	mouse β-*actin* 5' UTR, inverse sequence	This study
pRF-a9-P4-actin(inv)		This study
pRF-a9-P4(M11-M11)-actin(inv)	P4 mutagenesis series M1-M11	This study
pGL3-FLB-stop	based on pGL3-(EcoRV)	This study
pGL3-FLB-stop-TIE		This study
pGL3-FLB-stop-TIE-a9 IRES FL		This study
pGL3-FLB-stop-TIE-native		This study
pGL3-FLB-stop-TIE-P4-native		This study
pGL3-FLB-stop-TIE-P4(M5)-native		This study
pGL3-FLB-stop-TIE-HCV		This study
RNA transfection		
pcDNA3.1-5'UTR-3xHA-Nluc	46 nt scrambled UTR, kindly provided by C. Howard	(Osuna et al., 2017)
pcDNA3.1-actin(inv)-3xHA-Nluc		This study
pcDNA3.1-P4-actin(inv)-3xHA-Nluc		This study
pcDNA3.1-P4(M5)-actin(inv)-3xHA-Nluc		This study
CRISPR/Cas9 genome editing		
pX459-pSpCas9(BB)-2A-Puro-V2.0		Addgene #62988
pX459-a9(M5)-sgRNA-1		This study
Yeast		
rDNA constructs		
pNOY102	*URA3, 2μ, Gal7-rDNA-WT rRNA, amp*	(Nogi et al., 1991)
pNOY102-18S25Stag	*URA3, 2μ,Gal7-rDNA-tagged rRNA*	(Fujii et al., 2009)
pNOY102-18S25Stag-hES9S	*URA3, 2μ, Gal7-rDNA-tagged rRNA-hES9S*	This study
pNOY102-18S25Stag-hES9S-VA	*URA3, 2μ, Gal7-rDNA-tagged rRNA-hES9S-VA*	This study
pNOY102-18S25Stag-hES9S-VB	*URA3, 2μ, Gal7-rDNA-tagged rRNA-hES9S-VB*	This study
pNOY102-18S25Stag-hES9S-VC	*URA3, 2μ, Gal7-rDNA-tagged rRNA-hES9S-VC*	This study
pNOY373	*LEU2, 2μ, Pol1-rDNA-WT rRNA, amp*	(Nemoto et al., 2010)
pNOY373-18S25Stag	*LEU2, 2μ, Pol1-rDNA- tagged rRNA*	This study
pNOY373-18S25Stag-hES9S	*LEU2, 2μ, Pol1-rDNA- tagged rRNA-hES9S*	This study
pNOY373-18S25Stag-hES9S-VA	*LEU2, 2μ, Pol1-rDNA- tagged rRNA-hES9S-VA*	This study
pNOY373-18S25Stag-hES9S-VB	*LEU2, 2μ, Pol1-rDNA- tagged rRNA-hES9S-VB*	This study
pNOY373-18S25Stag-hES9S-VC	*LEU2, 2μ, Pol1-rDNA- tagged rRNA-hES9S-VC*	This study
Expression construct		
pRS316	*URA3, CEN*	Addgene #77145

Table S2: Yeast strains used in this study. Related to STAR Methods.

All yeast strains used and/or generated for this study are listed in the table.

Table S2. List of yeast strains

Strain	Genotype and Notes	Reference
NOY401	*MATA rpa190-3 ura3-1 leu2-3 trp1-1 can1-100*	(Nogi et al., 1991)
NOY401 WT rRNA	*MATA rpa190-3 ura3-1 leu2-3 trp1-1 can1-100 carrying pNOY102-WT rRNA::URA3*	This study
NOY401 tagged-hES9S	*MATA rpa190-3 ura3-1 leu2-3 trp1-1 can1-100 carrying pNOY102-tagged rRNA-hES9S::URA3*	This study
NOY401 tagged-hES9S-VA	*MATA rpa190-3 ura3-1 leu2-3 trp1-1 can1-100 carrying pNOY102-tagged rRNA-hES9S-VA::URA3*	This study
NOY401 tagged-hES9S-VB	*MATA rpa190-3 ura3-1 leu2-3 trp1-1 can1-100 carrying pNOY102-tagged rRNA-hES9S-VB::URA3*	This study
NOY401 tagged-hES9S-VB	*MATA rpa190-3 ura3-1 leu2-3 trp1-1 can1-100 carrying pNOY102-tagged rRNA-hES9S-VC::URA3*	This study
KAY488 (NOY890)	*MATA ura3-1 leu2-3,112 his3-11 trp1-1 ade2-1 can1-100 rdnaΔΔ::HIS3 carrying pRDN-hyg::URA3*	(Nemoto et al., 2010)
NOY890 WT rRNA	*MATA ura3-1 leu2-3,112 his3-11 trp1-1 ade2-1 can1-100 rdnaΔΔ::HIS3 carrying pNOY373-WT rRNA::LEU2*	This study
NOY890 tagged-hES9S	*MATA ura3-1 leu2-3,112 his3-11 trp1-1 ade2-1 can1-100 rdnaΔΔ::HIS3 carrying tagged pNOY373-rRNA-hES9S::LEU2*	This study
NOY890 tagged-hES9S-VA	*MATA ura3-1 leu2-3,112 his3-11 trp1-1 ade2-1 can1-100 rdnaΔΔ::HIS3 carrying tagged pNOY373-rRNA-hES9S-VA::LEU2*	This study
NOY890 tagged-hES9S-VB	*MATA ura3-1 leu2-3,112 his3-11 trp1-1 ade2-1 can1-100 rdnaΔΔ::HIS3 carrying tagged pNOY373-rRNA-hES9S-VB::LEU2*	This study
NOY890 tagged-hES9S-VC	*MATA ura3-1 leu2-3,112 his3-11 trp1-1 ade2-1 can1-100 rdnaΔΔ::HIS3 carrying tagged pNOY373-rRNA-hES9S-VC::LEU2*	This study

Table S3: DNA Oligonucleotides used in this study. Related to STAR Methods.

All DNA oligonucleotides used for cloning, RT-PCR, and RT-qPCR are listed in the table. F, forward primer; R, reverse primer.

Table S3. DNA oligonucleotides

Name	Sequence	Description
	qPCR primer	
KL050	TGGAGAATAACTTCTTCGTGGA	Rluc qPCR F
KL051	TTGGACGACGAACTTCACC	Rluc qPCR R
KL052	AAGAGATACGCCCTGGTTC	Fluc qPCR F
KL053	TTGTATTCAGCCCATATCGTTTC	Fluc qPCR R
KL056	GCCAACCGTGAAAAGATGAC	mouse β-*actin* F
KL057	CATCACAATGCCTGTGGTAC	mouse β-*actin* R
KL075	GAAGGCTCATGGCAAGAAGG	rabbit β-*globin* qPCR F
KL076	ATGATGAGACAGCACAATAACCAG	rabbit β-*globin* qPCR R
KL318	TGCAAACTCCTTGGTCACAC	y-UsnRNA1(SNR19) qPCR F
KL319	CAAACTTCTCCAGGCAGAAG	y-UsnRNA1(SNR19) qPCR R
KL320	CCATCATGAAGTGTGATGTC	y-*actin1* qPCR F
KL321	GACCTTCATGGAAGATGGAG	y-*actin1* qPCR R
KL412	CATGGCTGCAACACTTACACAGCA	mouse *NupL1* qPCR F
KL413	ATTGCAAGCCAGTGCCAATACCTG	mouse *NupL1* qPCR R
KL109	AAAAACAACCCAGCGAAGGC	mouse *Hoxa9* qPCR F
KL110	ATCGCTTCTTCCGAGTGGAG	mouse *Hoxa9* qPCR R
KL400	GGGAGCCGCGGTCTGA	mouse *Hoxa3* qPCR F
KL401	ACATGGAGGGAGCCATTTTTCA	mouse *Hoxa3* qPCR R
KL402	GAAGAAGATCCACGTGAGCG	mouse *Hoxa4* qPCR F
KL403	GGGTCAGGTAGCGGTTAAAGT	mouse *Hoxa4* qPCR R
KL404	GCGCAAGCTGCACATTAGTC	mouse *Hoxa5* qPCR F
KL405	TCAGGTAGCGGTTGAAGTGG	mouse *Hoxa5* qPCR R
KL406	GGCCACACTGAGGACAAGG	mouse *Hoxa11* qPCR F
KL407	GAACTCTCGCTCCAGCTCTC	mouse *Hoxa11* qPCR R
KL585	CCGTATGAAGGTCTGAGCGG	Nanoluc qPCR F
KL586	CAGTGTGCCATAGTGCAGGA	Nanoluc qPCR R
KL641	AGTGATTTACGCGTTATTGTTCTGCC	native qPCR F
KL642	TGTAACAACTTGGTGGCACCAG	native qPCR R
	qPCR primer for rRNA detection	
KL300	CTAGGCGAACAATGTTCTTAAAG	pre-mature 25S rRNA F
KL301	GACCTCAAATCAGGTAGGAGTACCC	mature 25S rRNA F
KL302	CACCGAAGGTACACTCGAGAGCTTC	tagged 25S rRNA R
KL303	CACCGAAGGTACCAGATTTC	endogenous 25S rRNA R
KL304	GCTTGTTGCTTCTTCTTTTAAGATAG	pre-mature 18S rRNA F
KL305	TACAGTGAAACTGCGAATGGC	mature 18S rRNA F
KL306	ATCTCTTCCAAAGGGTCGAG	endogenous 18S rRNA R
KL307	CGAGGATTCAGGCTTTGG	tagged 18S R
	PCR primer for rRNA strain characterization and ES9S sequencing	
KL314	GAACGAGACCTTAACCTACTAAATAGT	ES9S-span RT-PCR F
KL315	AAACCGATAGTCCCTCTAAGAAGT	ES9S-span RT-PCR R
KL316	GCTAATACATGCTTAAAATCTCGA	18Stag-span RT-PCR F
KL317	TTTTTATCTAATAAATACATCTCTTCCAA	18Stag-span RT-PCR R

KL394	GTGGTGCTAGCGCGG	ES9S-VC-span RT-PCR F
KL473	TCGATTCCGTGGGTGGTGG	18S rRNA-seq primer F
KL474	TAGCGCGCGTGCAGC	18S rRNA-seq primer R
	IRES confirmation in mouse embryos and RA-treated mESCs	
KL596	CTTCGTTGGCCACAATTAAAACAAACCAG	*a9* IRES primer (mouse) F
KL597	GCCCAGCAGGAAGGAGTC	*a9* CDS primer (mouse) R
	CRISPR/Cas9-editing of P4(M5) and genotyping	
KL598	**caccG**GGGCGGACACGTGACGCGCG	*a9*(M5) guide-1 F
KL599	**aaac**CGCGCGTCACGTGTCCGCCC**C**	*a9*(M5) guide-1 R
KL604	TTACGCGTTATTGTTCTGCCGGGCGGACACTATTCGCGCGTGGCCAATGGGGGCGCGGGCGCCG	30 nt overhang M5 template
KL605	CACCGGGCCATTAATAGCGTGCGGAGTGATTTACGCGTTATTGTTCTGCCGGGCGGACACTATTCGCGCGTGGCCAATGGGGGCGCGGGCGCCGGCAACTTATTAGGTGACTGTACTTCACCCC	60 nt overhang M5 template
KL596	CTTCGTTGGCCACAATTAAAACAAACCAG	*a9* IRES primer F (364 nt)
KL606	GCCGGGCGGACACTATT	M5-spec primer F (219 nt)
KL597	GCCCAGCAGGAAGGAGTC	*a9* CDS primer R
	In vitro transcription DNA template primer	
KL583	TCGAAATTAATACGACTCACTATAGGG	T7 IVT primer F
KL584	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTCTCGAGCGGC	polyA IVT primer R
KL589	TAATACGACTCACTATAGGGAGACTAGGCTTTTGCAAAAAGCTT	pGL3-T7 promoter F
KL588	TCTAGAATTACACGGCGATCTTTCCGCC	IVT_Fluc_R
	Hybrid ES9S sequences	
24 nt	CCTACTAAATAGT**GGTGCTAGCATTTGCTGGTTATCC**ACTTCTTAGAGG	Yeast WT ES9S
31 nt	CCTACTAAATAGT**TACGCGACCCCCGAGCGGTCGGCGTCCCCCA**ACTTCTTAGAGG	hES9S
33 nt	CCTACTAAATAGT**GGTGCTAGCCCCGAGCGGTCGGCCTGGTTATCC**ACTTCTTAGAGG	hES9S-VA
22 nt	CCTACTAAATAGT**GGTGCTAGGCGGCTGGTTATCC**ACTTCTTAGAGG	hES9S-VB
24 nt	CCTACTAAATAGT**GGTGCTAGCGCGGGCTGGTTATCC**ACTTCTTAGAGG	hES9S-VC