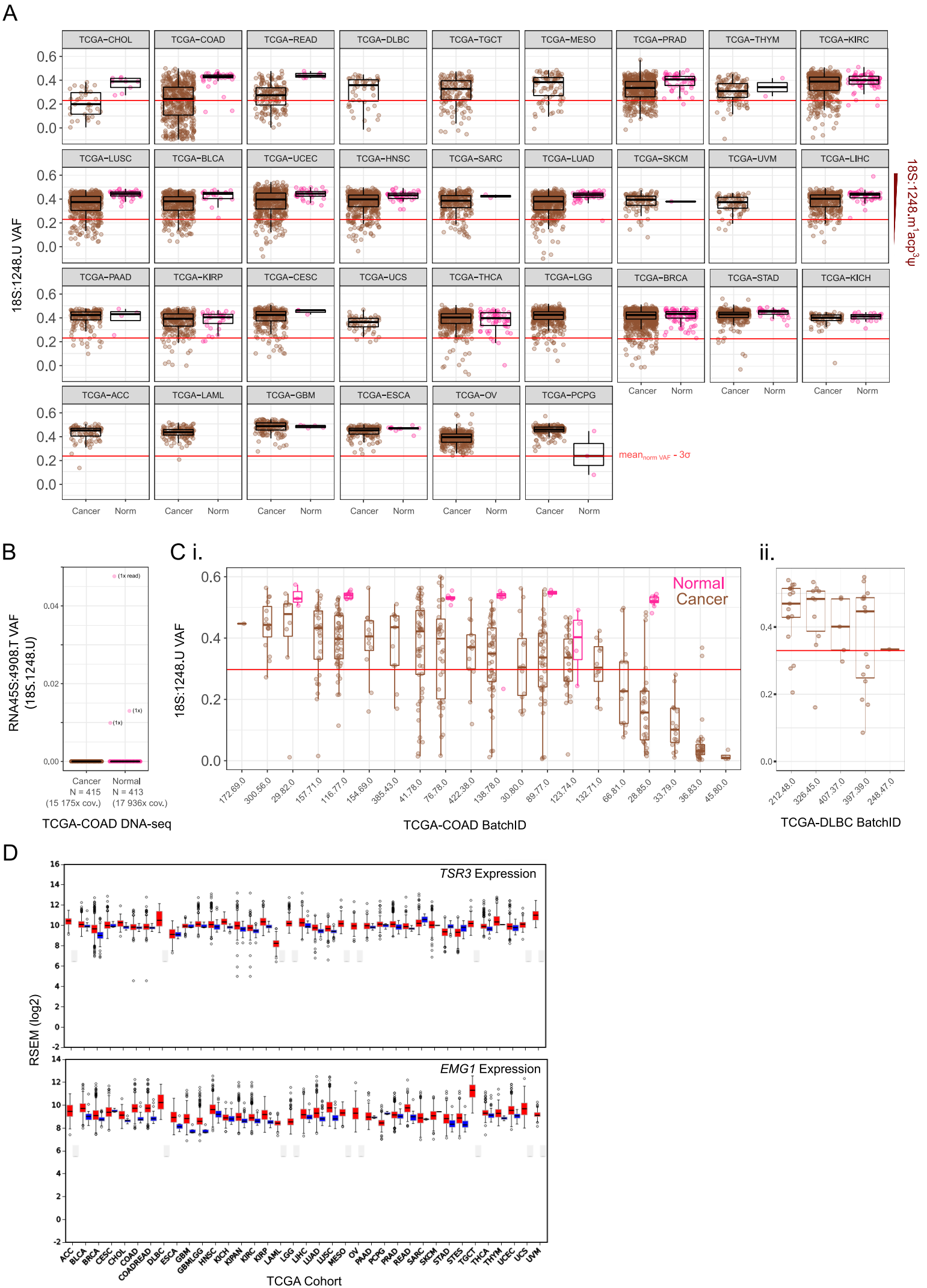
Cell Reports, Volume *31*

Supplemental Information

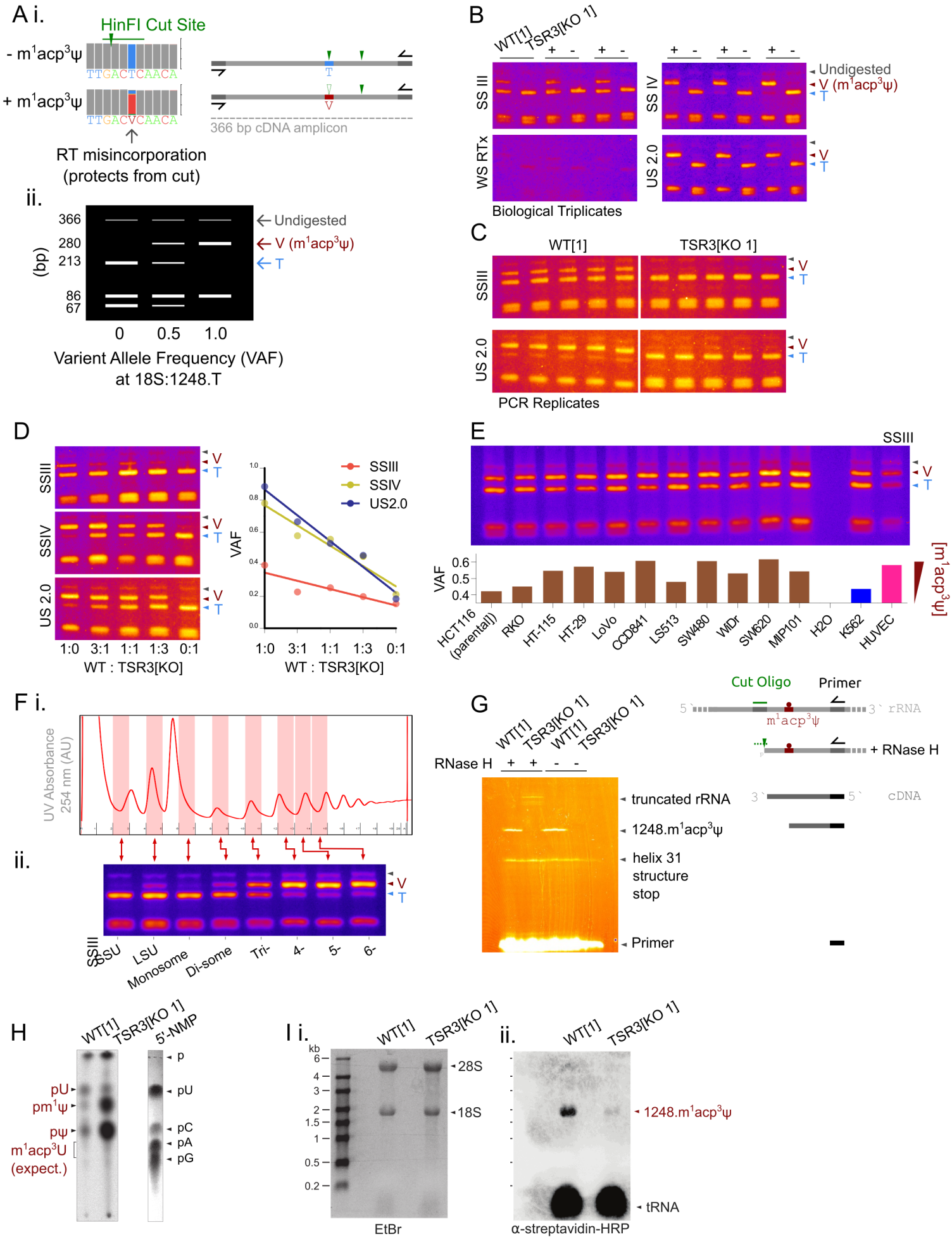
Loss of $m^1acp^3\Psi$ Ribosomal RNA Modification Is a Major Feature of Cancer

Artem Babaian, Katharina Rothe, Dylan Girodat, Igor Minia, Sara Djondovic, Miha Milek, Sandra E. Spencer Miko, Hans-Joachim Wieden, Markus Landthaler, Gregg B. Morin, and Dixie L. Mager


A
TCGA-CHOL
TCGA-COAD
TCGA-READ
TCGA-DLBC
TCGA-TGCT
TCGA-MESO
TCGA-PRAD
TCGA-THYM
TCGA-KIRC
TCGA-LUSC
TCGA-BLCA
TCGA-UCEC
TCGA-HNSC
TCGA-SARC
TCGA-LUAD
TCGA-SKCM
TCGA-UVM
TCGA-LIHC
TCGA-PAAD
TCGA-KIRP
TCGA-CESC
TCGA-UCS
TCGA-THCA
TCGA-LGG
TCGA-BRCA
TCGA-STAD
TCGA-KICH
TCGA-ACC
TCGA-LAML
TCGA-GBM
TCGA-ESCA
TCGA-OV
TCGA-PCPG
18S:1248.U VAF
18S:1248.m1acp3ψ
mean norm VAF - 3σ
Cancer
Norm
B
RNA45S:4908.T VAF (18S.1248.U)
(1x read)
(1x)
(1x)
Cancer N = 415 (15 175x cov.)
Normal N = 413 (17 936x cov.)
TCGA-COAD DNA-seq
C i.
18S:1248.U VAF
Normal
Cancer
172.69.0
300.56.0
29.82.0
157.71.0
116.77.0
154.69.0
385.43.0
41.78.0
76.78.0
422.38.0
138.78.0
30.80.0
89.77.0
123.74.0
132.71.0
66.81.0
28.85.0
33.79.0
36.83.0
45.80.0
TCGA-COAD BatchID
ii.
212.48.0
326.45.0
407.37.0
397.39.0
248.47.0
TCGA-DLBC BatchID
D
TSR3 Expression
EMG1 Expression
RSEM (log2)
ACC
BLCA
BRCA
CESC
CHOL
COAD
COADREAD
DLBC
ESCA
GBM
GBMLGG
HNSC
KICH
KIPAN
KIRC
KIRP
LAML
LGG
LIHC
LUAD
LUSC
MESO
OV
PAAD
PCPG
PRAD
READ
SARC
SKCM
STAD
STES
TGCT
THCA
THYM
UCEC
UCS
UVM
TCGA Cohort

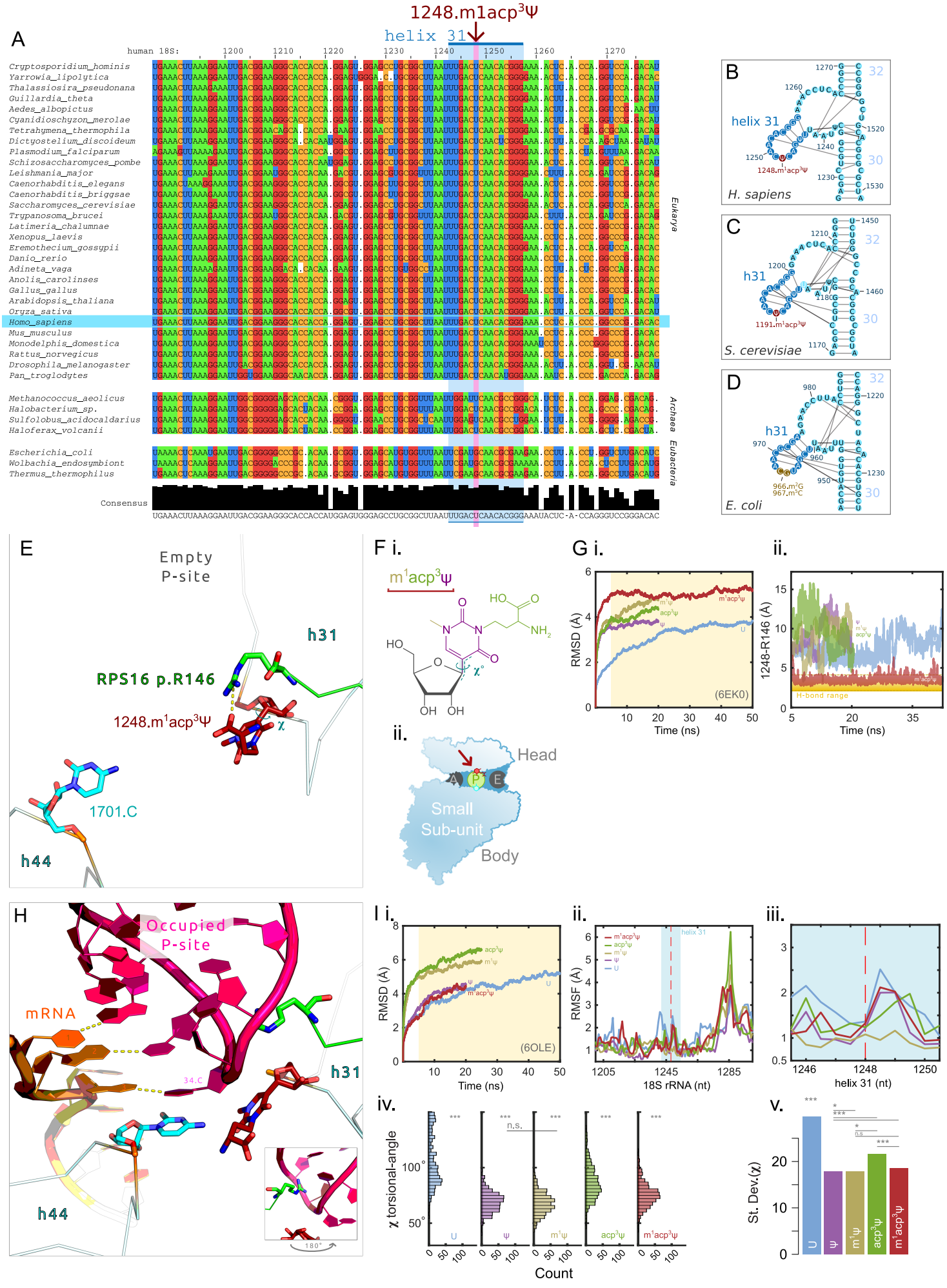
Supplementary Figure 1: Detailed look at hypo-m^{1}acp^{3}Ψ in the TCGA cohorts. Related to Figure 1.

A, The 18S.1248.U variant allele frequency (VAF) from 33 TCGA patient cohorts (study accessions in Table S1). **B**, The rDNA position underlying *18S.*1248.U (*RNA45S.*4908.T) is invariable in TCGA-COAD. Matched tumor-normal whole exome sequencing (WXS) data from 438 TCGA-COAD patients was aligned to *hgr1*. 415 cancer and 413 normal samples contained rDNA coverage of *RNA45S.*4908.T. Three normal samples had a single variant read each at this position, consistent with sequencing error. **C**, Batch-specific shift in the average 18S.1248.U VAF in i, TCGA-COAD and ii, TCGA-DLBC libraries. Similar to seen in the aRT-PCR assay for m^{1}acp^{3}Ψ (Fig. S2), there are batch-effects with m^{1}acp^{3}Ψ misincorporation, but the relative decrease in m^{1}acp^{3}Ψ-modification in CRC compared to normals is seen in across all batches. **D**, The gene expression of the m^{1}acp^{3}Ψ modifying enzymes *TSR3* and *EMG1* is not decreased or lost across the TCGA cohorts.


A i.
HinFI Cut Site
- m1acp3ψ
TTGACTCAACA
+ m1acp3ψ
TTGACVCAACA
T
V
366 bp cDNA amplicon
RT misincorporation
(protects from cut)
ii.
366
280
213
86
67
(bp)
Undigested
V (m1acp3ψ)
T
0
0.5
1.0
Varient Allele Frequency (VAF)
at 18S:1248.T
B
WT[1]
TSR3[KO 1]
SS III
SS IV
WS RTx
US 2.0
Undigested
V (m1acp3ψ)
T
Biological Triplicates
C
WT[1]
TSR3[KO 1]
SSIII
US 2.0
PCR Replicates
D
SSIII
SSIV
US 2.0
1:0
3:1
1:1
1:3
0:1
WT : TSR3[KO]
VAF
E
SSIII
VAF
HCT116 (parental)
RKO
HT-115
HT-29
LoVo
CCD841
LS513
SW480
WiDr
SW620
MIP101
H2O
K562
HUVEC
[m1acp3ψ]
F i.
UV Absorbance
254 nm (AU)
ii.
SSIII
SSU
LSU
Monosome
Di-some
Tri-
4-
5-
6-
G
WT[1]
TSR3[KO 1]
WT[1]
TSR3[KO 1]
RNase H
Cut Oligo
Primer
rRNA
+ RNase H
cDNA
truncated rRNA
1248.m1acp3ψ
helix 31
structure
stop
Primer
H
WT[1]
TSR3[KO 1]
5'-NMP
pU
pm1ψ
pψ
m1acp3U
(expect.)
p
pU
pC
pA
pG
I i.
WT[1]
TSR3[KO 1]
kb
6
4
3
2
1.5
1
0.5
0.2
28S
18S
EtBr
ii.
WT[1]
TSR3[KO 1]
1248.m1acp3ψ
tRNA
α-streptavidin-HRP

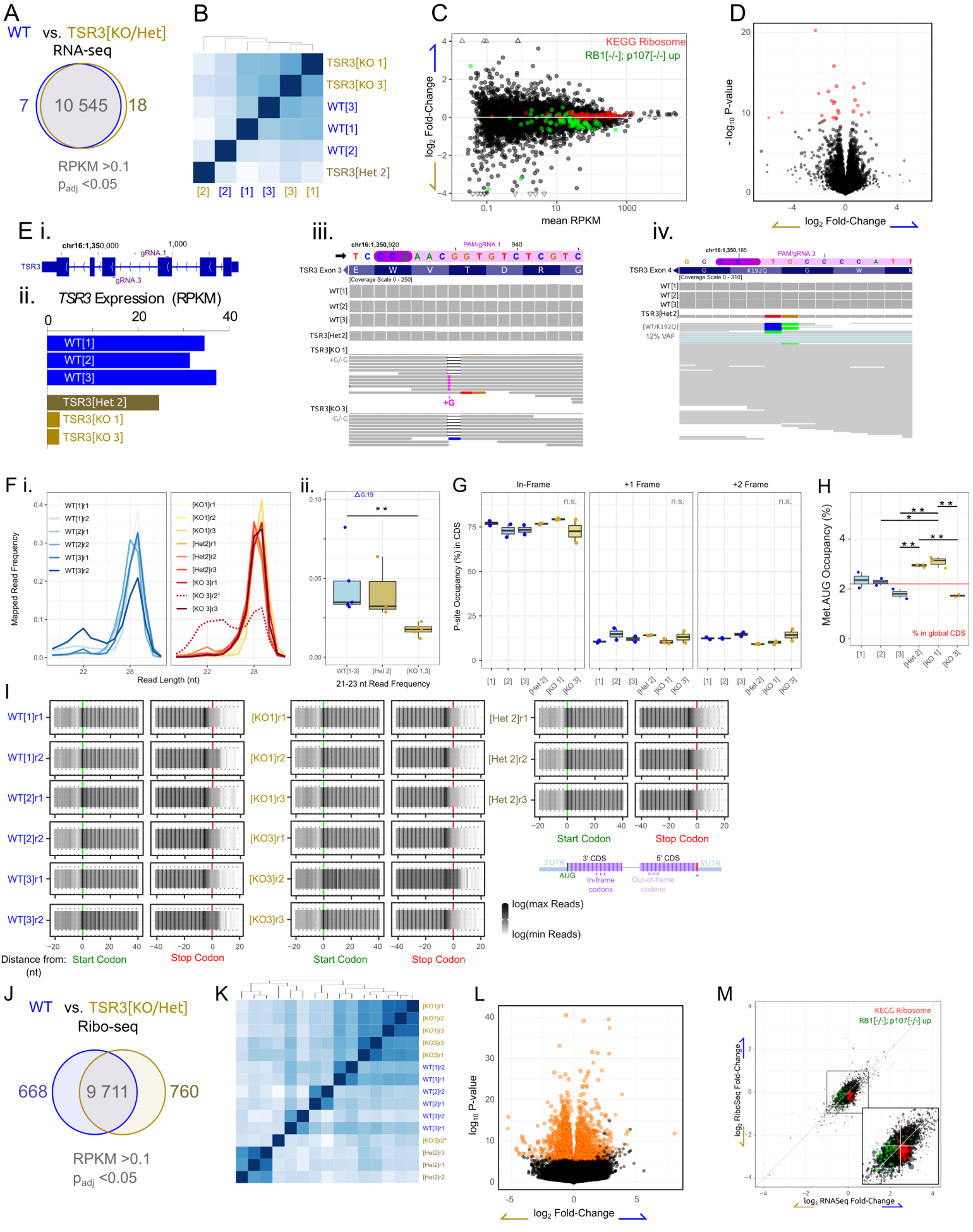
Supplementary Figure 2: Aminocarboxyl propyl Reverse Transcription (aRT)-PCR assay for $m^1acp^3\Psi$ and rRNA modification in cell lines. Related to Figure 3 and STAR Methods.

A i, The 18S.1248.$m^1acp^3\Psi$ modification assay is based on the misincorporation of nucleotides in first strand complementary DNA (cDNA) strand synthesis by reverse transcriptase (RT). The cDNA is then PCR amplified and **ii,** the ratio of reference T and not-T (V = A, C, or G) is genotyped by the *HinFI* restriction enzyme cut-site which overlaps 18S.1248. **B,** The choice of RT-enzyme; SuperScript III (SSIII), SuperScript IV (SSIV), WarmStart RTx (WS RTx) or, UltraScript 2.0 (US 2.0), influences nucleotide misincorporation rates and the variant allele frequency (VAF) read-out of the m^1acp^3-assay, although VAF remains consistent across biological replicates of input RNA of the colorectal cancer (CRC) cell line HCT116 wildtype clone 1 (WT[1]), or HCT116 with *TSR3* gene knockout clone 1 (*TSR3*[KO 1]). **C,** PCR replicates of WT[1] and TSR3[KO 1] cDNA, shows consistent readout. **D,** HCT116 WT[1] and *TSR3*[KO 1] RNA was mixed at fixed weight ratios (µg total RNA) prior to RT to determine if the assay is quantitative for m_1acp_3 modification. **E,** The aRT-PCR assay applied to 11 CRC cell lines, primary human umbilical vein endothelial cells (HUVEC) as a normal control and, the blast-phase chronic myelogeous leukemia cell line K562 as a hypo-$m^1acp^3\Psi$ positive control. **F i,** Polysomal fractionation and **ii,** sub-fraction $m^1acp^3\Psi$ RT-PCR assay of the hypo-$m^1acp^3\Psi$ cell line K562. In cells containing a mixture of +/- $m^1acp^3\Psi$ modification, unmodified rRNA incorporates into mature ribosomes and is enriched in the lower-order mono- and di-somes. **G,** Primer extension assay for 18S.1248.$m^1acp^3\Psi$ modification in HCT116 WT[1] and *TSR3*[KO 1]. The helix 31 structural stop and rRNA truncation via DNA cut oligo + RNase H treatment is used as internal load controls. **H,** SCARLET assay of HCT116 WT[1] and *TSR3*[KO 1] is not suitable for measuring $m^1acp^3\Psi$ as this base may interfere with probe hybridization, but precursor bases were detected and accumulated in the TSR3 knock-out cells. **I,** Total RNA treated with the amino-reactive ester, N-hydroxysuccinimidobiotin and detected with **i.** ethidium bromide staining as load control and **i.** anti-streptavidin-HRP antibody and. *TSR3*[KO 1] cells show loss of primary amine reactivity (such as $m^1\underline{a}cp^3\Psi$) in 18S rRNA but not in tRNA.


A
1248.m1acp3Ψ
helix 31
human 18S:
1200
1210
1220
1230
1240
1250
1260
1270
Cryptosporidium_hominis
Yarrowia_lipolytica
Thalassiosira_pseudonana
Guillardia_theta
Aedes_albopictus
Cyanidioschyzon_merolae
Tetrahymena_thermophila
Dictyostelium_discoideum
Plasmodium_falciparum
Schizosaccharomyces_pombe
Leishmania_major
Caenorhabditis_elegans
Caenorhabditis_briggsae
Saccharomyces_cerevisiae
Trypanosoma_brucei
Latimeria_chalumnae
Xenopus_laevis
Eremothecium_gossypii
Danio_rerio
Adineta_vaga
Anolis_carolinses
Gallus_gallus
Arabidopsis_thaliana
Oryza_sativa
Homo_sapiens
Mus_musculus
Monodelphis_domestica
Rattus_norvegicus
Drosophila_melanogaster
Pan_troglodytes
Methanococcus_aeolicus
Halobacterium_sp.
Sulfolobus_acidocaldarius
Haloferax_volcanii
Escherichia_coli
Wolbachia_endosymbiont
Thermus_thermophilus
Eukarya
Archaea
Eubacteria
Consensus
B
helix 31
H. sapiens
C
h31
S. cerevisiae
D
h31
E. coli
E
Empty
P-site
h31
RPS16 p.R146
1248.m1acp3Ψ
χ
1701.C
h44
F i.
m1acp3Ψ
ii.
Head
Small
Sub-unit
Body
G i.
RMSD (Å)
Time (ns)
(6EK0)
ii.
1248-R146 (Å)
H-bond range
H
Occupied
P-site
mRNA
34.C
h31
h44
180°
I i.
RMSD (Å)
Time (ns)
(6OLE)
ii.
RMSF (Å)
helix 31
18S rRNA (nt)
iii.
helix 31 (nt)
iv.
χ torsional-angle
n.s.
Count
v.
St. Dev.(χ)

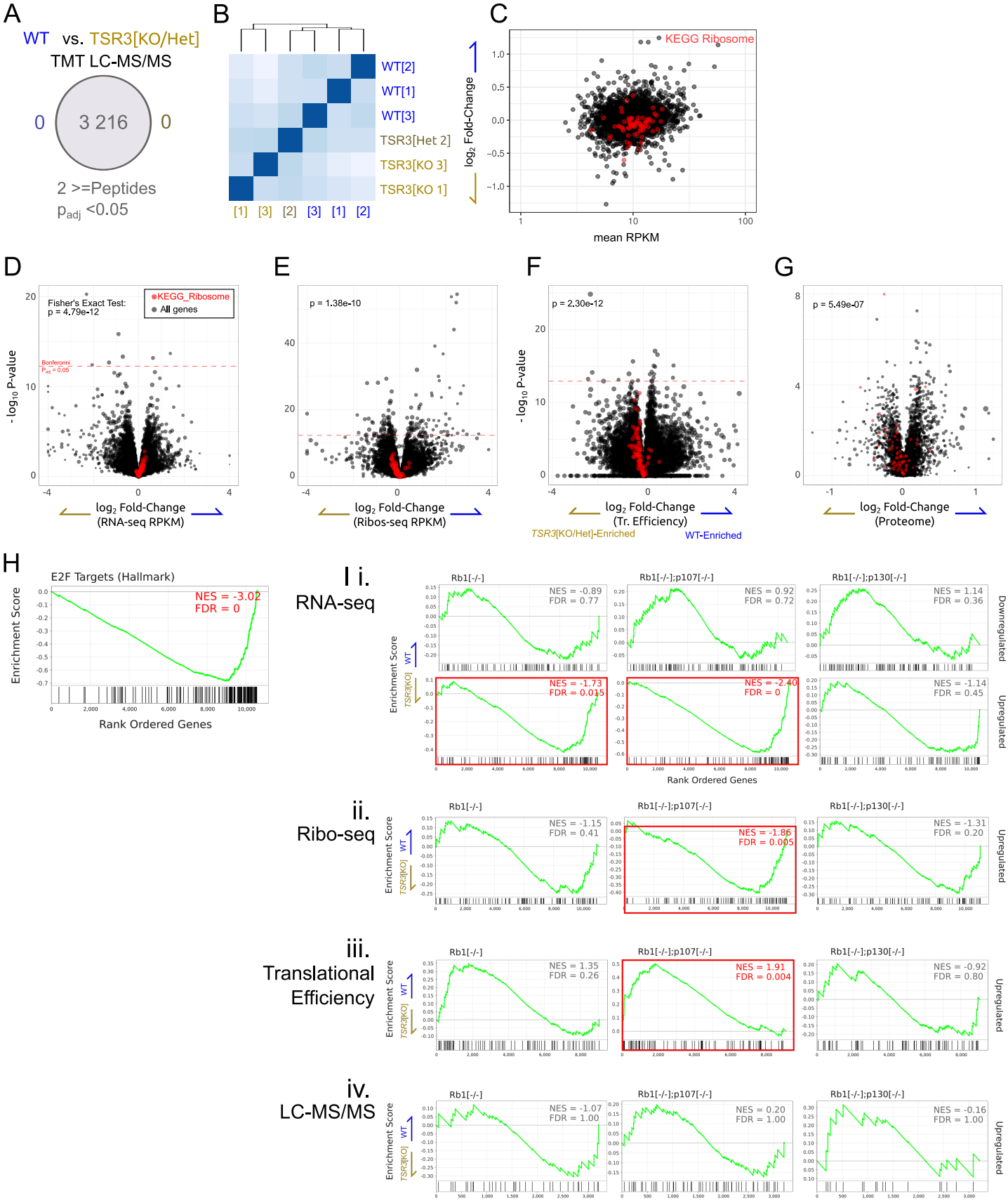
Supplementary Figure 3: Evolutionary and structural analysis of 18S.1248.m^{1}acp$^3\Psi$. Related to Figure 2.

18S.1248.m^{1}acp$^3\Psi$ is absolutely conserved in the *Eukarya* peptidyl decoding site. **A,** Evolutionary conservation of 18S rRNA in the locus surrounding helix 31 in *Eukarya* and select *Archaea* and *Eubacteria* species (Bernier et al., 2014; Petrov et al., 2014). Burgendy arrow indicates the position homologous to human18S:1248.U. **B-D,** The conserved secondary structure of helix 31 and its known modification sites (Machnicka et al., 2013; Petrov et al., 2014). **E-I,** Whole ribosome molecular dynamics simulations (MD) were ran for up to 50ns with 18S.1248.m^{1}acp$^3\Psi$, acp$^3\Psi$, m$^1\Psi$, Ψ or U base in an empty (**E-G**, PDB: 6EK0) or tRNA-occupied P-site (**H,I**, PDB: 6OLE). **F,** The m^{1}acp$^3\Psi$ modification stabilizes the decoding peptidyl (P-) site via a hydrogen bond with the universally conserved RPS16 (uS9) p.R146 residue. **E,** 20ns simulation showing bonding between 18S.1248.m^{1}acp$^3\Psi$ and RPS16 p.R146 compared to **H,** the structure with mRNA and P-site tRNA. **G,i.** Root mean squared deviation (RMSD) of empty P-site MD atoms indicate simulations stabilize after 5ns (with up to ~6Å RMSD), up to 45ns (yellow highlight) was used for analysis. **ii,** Minimal distance between m^{1}acp$^3\Psi$ (3-carboxyl oxygen) or uridine (4-oxygen) and the closest guanidinium hydrogen of RPS16 p.R146 supports that m^1 and acp^3 are necessary to be within 3.5Å for hydrogen bonding and nucleotide stabilization. **I,** RMSD of P-site occupied simulations with **ii.** Root mean squared fluctuation (RMSF) of 18S rRNA adjacent to 18S.1248 and **iii,** within helix 31 shows 18S.1248.m^{1}acp$^3\Psi$ is comparable relative to other modification states. The average (**iv**) and standard deviation of χ-angle of the 18S.1248 nucleotide in each simulation shows acp^3 increases torsional variation of the base which is stabilized by the m^1 modification. We postulate that 18S.1248 acp^3-modification is involved in coordinating RPS16 p.146 for P-site tRNA positioning and contributes to the stability of the decoding core. Statistical difference between χ-angle mean (Tukey HSD), and standard deviation (Benjamini-Hochberg corrected F-tests) were used (* is $p < 0.05$, ** is $p < 0.001$ and *** is $p < 0.0001$).


A
WT vs. TSR3[KO/Het]
RNA-seq
7
10 545
18
RPKM >0.1
padj <0.05
B
TSR3[KO 1]
TSR3[KO 3]
WT[3]
WT[1]
WT[2]
TSR3[Het 2]
[2] [2] [1] [3] [3] [1]
C
KEGG Ribosome
RB1[-/-]; p107[-/-] up
log2 Fold-Change
mean RPKM
D
- log10 P-value
log2 Fold-Change
E i.
chr16:1,350,000
gRNA.1
gRNA.3
TSR3
ii.
TSR3 Expression (RPKM)
WT[1]
WT[2]
WT[3]
TSR3[Het 2]
TSR3[KO 1]
TSR3[KO 3]
iii.
chr16:1,350,920
PAM/gRNA.1
TSR3 Exon 3
[Coverage Scale 0 - 250]
+G
iv.
chr16:1,350,185
PAM/gRNA.3
TSR3 Exon 4
[Coverage Scale 0 - 310]
K192Q
[WT/K192Q]
12% VAF
F i.
Mapped Read Frequency
Read Length (nt)
ii.
21-23 nt Read Frequency
G
In-Frame
+1 Frame
+2 Frame
n.s.
P-site Occupancy (%) in CDS
H
Met.AUG Occupancy (%)
% in global CDS
I
Start Codon
Stop Codon
3'UTR
3' CDS
5' CDS
5'UTR
AUG
In-frame codons
Out-of-frame codons
log(max Reads)
log(min Reads)
Distance from: Start Codon Stop Codon (nt)
J
WT vs. TSR3[KO/Het]
Ribo-seq
668
9 711
760
RPKM >0.1
padj <0.05
K
L
log10 P-value
log2 Fold-Change
M
KEGG Ribosome
RB1[-/-]; p107[-/-] up
log2 RiboSeq Fold-Change
log2 RNASeq Fold-Change

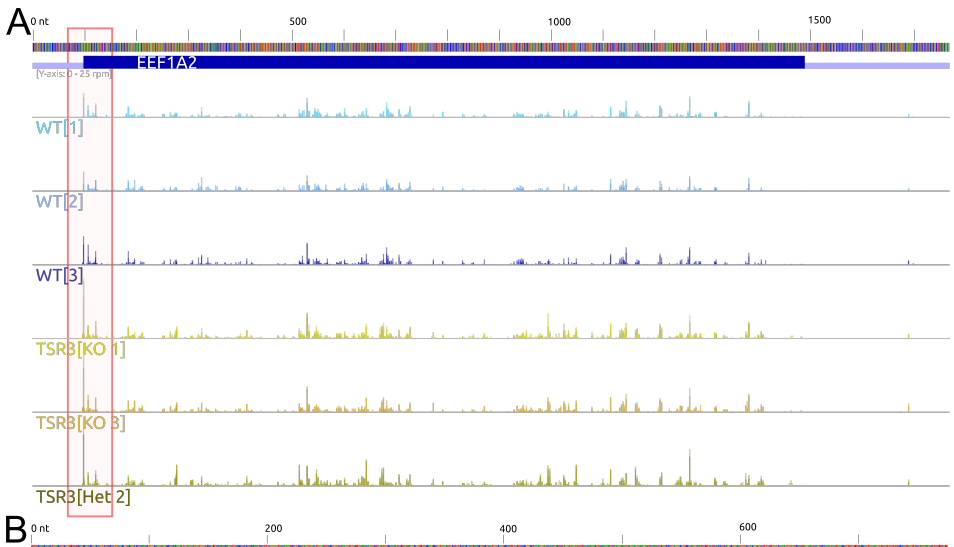
Supplementary Figure 4: RNA-seq and Ribo-seq of HCT116 *TSR3*[KO/Het] cells. Related to Figure 3.

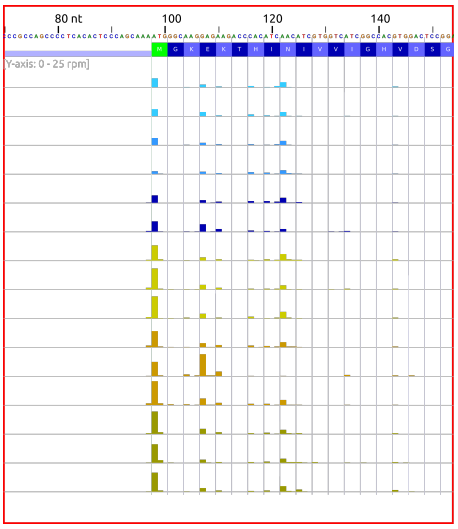
HCT116 WT[1-3] versus *TSR3*[KO]/[Het] RNA-seq and Ribo-seq metrics. **A,** Differential mRNA expression of expressed (reads per million kilobase, $RPKM_{mRNA}$ >0.1) between WT[1-3] and *TSR3*[KO]/[Het] clones. **B,** Hierarchical clustering of libraries based on expressed genes. Globally, *TSR3*[Het 2] is more dissimilar to *TSR3*[KO 1,2] clones. **C,** MA-plot for mRNA expression highlighting genes in the '*KEGG_RIBOSOME'* and '*RB_P107_DN.V1_UP*' gene sets (see: Fig 3). **D,** Volcano plot for differentially expressed genes between WT and *TSR3*[KO]/[Het] with statistically significant genes highlighted in red (list is available in Table S1C). **E,i.** *TSR3* gene structure and **ii.** decreased expression in *TSR3*[KO]/[Het] lines measured by RNA-seq (in reads per kilobase per million mapped reads, RPKM). IGV screenshots showing the CRISPR-Cas9 editing sites in (**iii.**) *TSR3*[KO 1,3], and (**iv.**) *TSR3*[Het 2]. **F-M,** HCT116 WT[1-3] versus *TSR3*[KO]/[Het] ribo-seq metrics. **F, i,** As a quality control metric, the length distribution of mapped ribosome-protected fragments for each of the two WT[1-3], and three *TSR3*[KO]/[Het] biological replicates of ribo-seq libraries were plotted. The *TSR3*[KO 3] biological replicate 2 (r2) library had a bi-modal read-length distribution, peaking at 22 and 28 nt suggesting incomplete cycloheximide treatment (Lareau et al., 2014) thus, this library was excluded from downstream expression and positional analyses. **ii,** Short (21-23 nt) ribosome fragments coincide with ribosomes stalled in the rotated, post peptide-bond state (Lareau et al., 2014). *TSR3*[KO 1,3] libraries had fewer short fragments implying $m^1acp^3\Psi$-deficient ribosomes have a lower probability of being in the rotated transition state relative to WT ribosomes. **G,** The P-site periodicity within each library showed the majority of CDS ribosomes were in-frame, with no significant difference in frame-shifting upon $m^1acp^3\Psi$ perturbation. **H,** P-site occupancy was calculated over all expressed coding sequences (CDS). Globally, there was no significant difference between P-site occupancy per codon in WT[1-3] in *TSR3*[KO]/[Het] libraries. Since 18S.1248.$m^1acp^3\Psi$ is located at the P-site where initiation codon selection occurs, we tested if the initiation AUG codon was differentially occupied between any genotypes. *TSR3*[Het 2] and *TSR3*[KO 1], but not *TSR3*[KO 3] have elevated AUG occupancy relative to WT. **I,** Metagene heatmap of P-site occupancy around start and stop codons. For each read optimal P-site offset was determined (13 nt from 5' end) and P-site occupancy computed. Log10-transformed P-site coverage is shown for all replicates, scaled by min and max number of reads per sample. A clear trinucleotide periodicity in the CDS of transcripts is observed. **J,** Hierarchical clustering of libraries based on total translation recapitulates mRNA clustering. **D,** Differential translation of ribo-seq expressed ($RPKM_{Ribo}$ > 0.1) genes. **F,** Volcano-plot for total translation, highlighting the ribosome and RB/E2F gene sets. **M,** The $\log_2$ mRNA fold-change and $\log_2$ translation fold-change (WT / *TSR3*[KO]/[Het]) and detail inlay, with each gene size-scaled by mean $RPKM_{mRNA}$. RB/E2F gene sets are highlighted, showing RP genes are more efficiently translated in *TSR3*[KO]/[Het] clones (all points below diagonal in, see also Fig 3).. Tukey HSD test was used for testing a statistical difference between group means (* is $p < 0.05$, ** is $p < 0.001$ and *** is $p < 0.0001$).

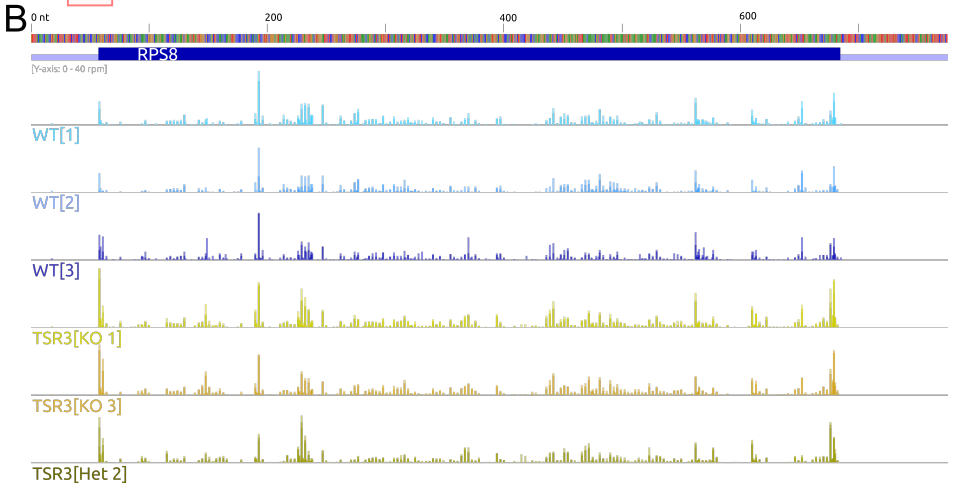

A
WT vs. TSR3[KO/Het]
TMT LC-MS/MS
0
3 216
0
2 >=Peptides
p_adj <0.05
B
WT[2]
WT[1]
WT[3]
TSR3[Het 2]
TSR3[KO 3]
TSR3[KO 1]
[1] [3] [2] [3] [1] [2]
C
KEGG Ribosome
log2 Fold-Change
mean RPKM
D
Fisher's Exact Test:
p = 4.79e-12
KEGG_Ribosome
All genes
Bonferonni
- log10 P-value
log2 Fold-Change
(RNA-seq RPKM)
E
p = 1.38e-10
log2 Fold-Change
(Ribos-seq RPKM)
F
p = 2.30e-12
- log10 P-value
log2 Fold-Change
(Tr. Efficiency)
TSR3[KO/Het]-Enriched
WT-Enriched
G
p = 5.49e-07
log2 Fold-Change
(Proteome)
H
E2F Targets (Hallmark)
NES = -3.02
FDR = 0
Enrichment Score
Rank Ordered Genes
I i.
RNA-seq
ii.
Ribo-seq
iii.
Translational
Efficiency
iv.
LC-MS/MS
Rb1[-/-]
Rb1[-/-];p107[-/-]
Rb1[-/-];p130[-/-]
NES = -0.89
FDR = 0.77
NES = 0.92
FDR = 0.72
NES = 1.14
FDR = 0.36
NES = -1.73
FDR = 0.015
NES = -2.40
FDR = 0
NES = -1.14
FDR = 0.45
NES = -1.15
FDR = 0.41
NES = -1.86
FDR = 0.005
NES = -1.31
FDR = 0.20
NES = 1.35
FDR = 0.26
NES = 1.91
FDR = 0.004
NES = -0.92
FDR = 0.80
NES = -1.07
FDR = 1.00
NES = 0.20
FDR = 1.00
NES = -0.16
FDR = 1.00
Downregulated
Upregulated
WT
TSR3[KO]

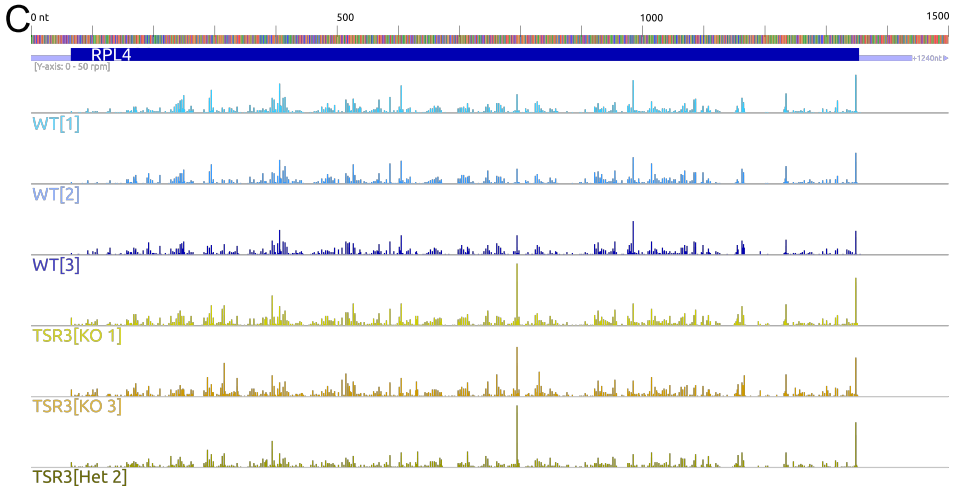
Supplementary Figure 5: Liquid-Chromatography Tandem Mass Spetrometry (LC-MS/MS) validation and RB/E2F transcriptional signatures. Related to Figure 3.

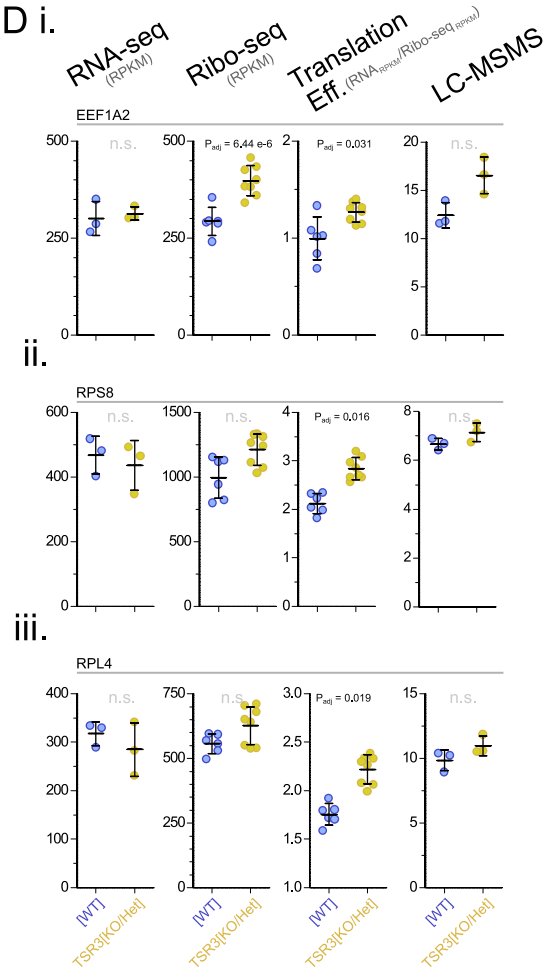
HCT116 WT[1-3] versus *TSR3*[KO]/[Het] LC-MS/MS results. **A,** After multiple-testing correction, no proteins were significantly different between genotypes, and no TSR3 peptides were found in the samples. **D-G**, Volcano plots for RNA-seq, ribo-seq, translational efficiency ($\text{Ribo-seq}_{RPKM}/\text{RNA-seq}_{RPKM}$) and proteomic data with genes in the `KEGG_RIBOSOME` gene-set highlighted. As a complimentary analysis to GSEA in Figure 3F,H, a Fisher's Exact Test was performed to test for enrichment of `KEGG_RIBOSOME` genes either up- or down-regulated between genotypes, compared to all other genes. **H,** HCT116 WT[1-3] versus *TSR3*[KO]/[Het] Gene Set Enrichment analysis (GSEA) for the E2F Targets and oncogenic signature and **I,** *RB1*, *RBL1* (p107) knock-out gene signatures. **i.** Gene set mRNA expression is enriched specifically in genes upregulated upon Rb1 and Rb1;p107 knockout (gene sets: *RB_DN.V1_UP*, *RB_P107_DN.V1_UP*). **ii.** The translational output (ribo-seq signal) of genes upregulated in Rb1;p107 knockout remains increased but **iii.** this gene set is translated less efficiently and **iv.** has no difference at the proteome level.

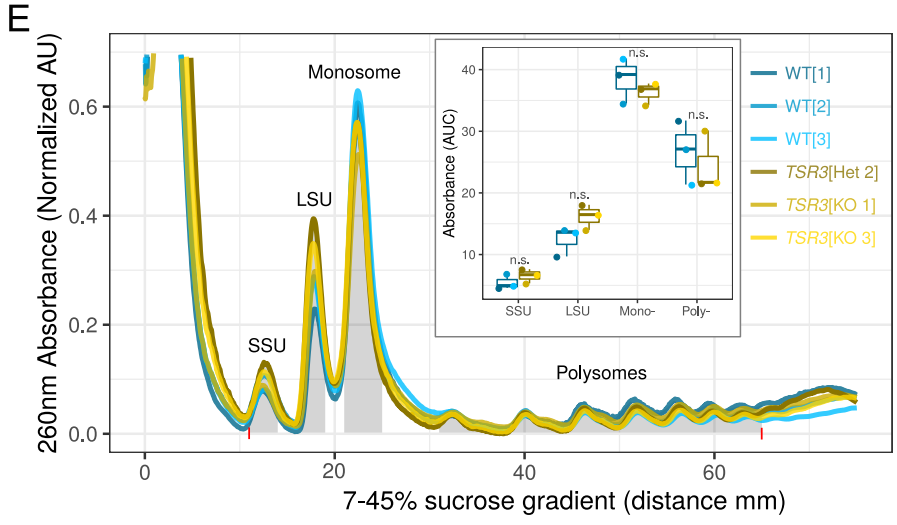

A
EEF1A2
WT[1]
WT[2]
WT[3]
TSR3[KO 1]
TSR3[KO 3]
TSR3[Het 2]


80 nt
100
120
140


B
RPS8
WT[1]
WT[2]
WT[3]
TSR3[KO 1]
TSR3[KO 3]
TSR3[Het 2]


C
RPL4
WT[1]
WT[2]
WT[3]
TSR3[KO 1]
TSR3[KO 3]
TSR3[Het 2]


D i.
RNA-seq (RPKM)
Ribo-seq (RPKM)
Translation Eff. (RNA RPKM/Ribo-seq RPKM)
LC-MSMS
EEF1A2
n.s.
P adj = 6.44 e-6
P adj = 0.031
ii.
RPS8
P adj = 0.016
iii.
RPL4
P adj = 0.019
[WT]
TSR3[KO/Het]


E
260nm Absorbance (Normalized AU)
Monosome
LSU
SSU
Polysomes
7-45% sucrose gradient (distance mm)
Absorbance (AUC)
SSU
LSU
Mono-
Poly-
WT[1]
WT[2]
WT[3]
TSR3[Het 2]
TSR3[KO 1]
TSR3[KO 3]

Supplementary Figure 6: Representative changes to translational efficiency. Related to Figure 3.

Ribo-seq traces for representative genes **A**, *EEF1A2* (eukaryotic translation elongation factor 1 alpha 2), **B**, *RPS8* (ribosomal protein S8), and **C**, *RPL4* (ribosomal protein L4) for HCT116 WT[1-3] versus *TSR3*[KO]/[Het]. **D**, Matched RNA expression (in reads per kilobase per million mapped reads, RPKM), Ribo-seq CDS expression (RPKM), translational efficiency ($\text{Ribo-seq}_{\text{RPKM}}/\text{RNA-seq}_{\text{RPKM}}$), and proteomic expression (median normalized expression values). **F**, Polysome profiles and relative area under the curve (AUC) intensity for sub-fractions (highlighted in gray) (inlay). Total absorbance was normalized to 100 artificial units in the working range of the gradient (between red ticks on x-axis).