Loss of $m^1acp^3\Psi$ ribosomal RNA modification is a major feature of cancer

Artem Babaian[1,2]*, Katharina Rothe[1,2], Dylan Girodat[3], Igor Minia[4], Sara Djondovic[1], Miha Milek[4], Hans-Joachim Wieden[3], Markus Landthaler[4,5], Gregg Morin[1,6], and Dixie L. Mager[1,2].

1. Department of Medical Genetics, University of British Columbia, Vancouver, BC, Canada

2. Terry Fox Laboratory, BC Cancer, Vancouver, BC, Canada

3. Alberta RNA Research and Training Institute, University of Lethbridge, Lethbridge, AB, Canada

4. Max Delbrück Center for Molecular Medicine Berlin in the Helmholtz Association, Berlin Institute for Medical Systems Biology, Berlin, Germany

5. IRI Life Sciences, Institut für Biologie, Humbold Universität, Berlin, Germany

6. Michael Smith Genome Sciences Centre, British Columbia Cancer Agency, Vancouver, BC, Canada

Words: **1,591**

***** Corresponding author

Dr. Artem Babaian, Ph.D.
Terry Fox Laboratory
BC Cancer
675 West 10th Avenue
Vancouver, BC, V5Z1L3, Canada
Email: ababaian@bccrc.ca

Abstract

The ribosome is an RNA-protein complex essential for translation in all domains of life. The structural and catalytic core of the ribosome is its ribosomal RNA (rRNA). While mutations in ribosomal protein (RP) gene are known drivers of oncogenesis, oncogenic rRNA variants have remained elusive.

We discovered a cancer-specific single nucleotide variation at 18S.1248U in the 18S rRNA of up to 45.9% colorectal carcinoma (CRC) patients and across >22 cancer types. This is the site of a unique hyper-modified base, 1-methyl-3-α-amino-α-carboxyl-propyl pseudouridine ($m^1acp^3\Psi$), a modification that is >1 billion years conserved at the ribosome's peptidyl decoding-site. A sub-set of CRC tumors we term 'hypo-$m^1acp^3\Psi$', show sub-stoichiometric $m^1acp^3\Psi$-modification unlike normal control tissues. Our $m^1acp^3\Psi$ knockout model and hypo-$m^1acp^3\Psi$ patient tumors share a translational signature, characterized by highly abundant ribosomal proteins.

Thus, $m^1acp^3\Psi$-deficient rRNA forms an uncharacterized class of 'onco-ribosome' which may serve as an innovative chemotherapeutic target for treating cancer patients.

Introduction

The ribosome is a massive ribonucleoprotein particle (RNP) responsible for the transformation of genetic information encoded as nucleic acids into functional proteins encoded as amino acids. Unlike most RNPs, it is ribosomal RNA (rRNA), and not ribosomal proteins (RPs) that form the most ancient and catalytic core of the complex[1]. rRNA is further functionalized by a constellation of at least 14 distinct chemical modifications across 200+ sites,[2] clustering around active sites of the ribosome[3], yet the function of many rRNA modifications remain unclear.

The human ribosome contains >80 RPs and four rRNAs, totaling ~80% of cellular RNA. During the initial human genome sequencing project, ribosomal DNA (rDNA) loci were systematically excluded from the reference genome[4]; given that a reference sequence of the rRNA gene, *RNA45S*, was available and the 80-800 rDNA copies were believed to be homogeneous[5], although there was early evidence for rDNA polymorphism in humans[6,7]. Thus, as technology and sequencing consortium projects revolutionized genomics and transcriptomics, our understanding of rDNA variation has lagged.

rDNA sequence variation at the intra- and inter-individual level has been documented in multiple species including humans[8–12], but the functional implications of rDNA variation remain elusive. Mutation of RP genes and ribosome biogenesis factors can cause a class of diseases termed ribosomopathies, including Diamond Blackfan anemia (DBA)[13], and some cancers[14]. It has been hypothesized that cancer cells contain a functionally specialized class of ribosomes to facilitate rapid protein synthesis, termed the “onco-ribosomes”[15,16]. Cancer genomics has supported this notion with the identification of several oncogenic driver mutations in RP genes[14], the best characterized of which are RPL10 (uL16) p.R98S in T-cell acute lymphoblastic leukemia[17,18] and RPS15 (uS19) C’ terminal mutations in chronic lymphocytic leukemia[19]. In addition, germline mutations such as in DBA patients and in *RPS20* can cause heredity cancers including colorectal carcinoma (CRC)[20,21].

As RP mutations have been implicated in tumorigenesis, we hypothesized that rRNA variation or mutation is a cancer driver. To map functional rRNA sequence variation, we considered tumorigenesis as a natural experiment in which polymorphic and mutant rRNA alleles undergo selective evolutionary change in frequency within each patient. We discovered a surprising 18S rRNA single nucleotide variation at the decoding core of the ribosomal peptidyl (P)-site, affecting up to 45.9% of CRC patients, making this the most frequent ribosomal variant associated with cancer to date and potentially revolutionizing future chemotherapeutic strategies against this disease.

Results & Discussion

An unexpected rRNA variant in cancer: sub-stoichiometric modification of 18S.1248.m^{1}acp^{3}Ψ

In an initial screen for cancer-driver rRNA variants, we aligned RNA-seq reads from 66 colorectal carcinoma (CRC) tumors and patient-matched adjacent normal tissue to a single-copy reference rDNA. To test for allelic selection inconsistent with neutral drift, the patient-matched difference in expressed variant allele frequency (VAF) was measured for deviation from zero for each position of 18S and 28S (Fig. 1a). A single nucleotide variation deviated from neutrality, 18S:r.1248.U (p_{adj} = 3.81e-8). The 18S.1248.U variation is recurrent selection over non-U or 18S.1248V alleles in a striking 44.9% of CRC patients (Fig. 1a); in comparison, oncogenic *KRAS*^{G12} codon mutation occurs in only 36% of CRC patients[22].

Surprisingly, at the DNA level, *RNA45S*:1248.T is invariable in humans[10] and in the mature rRNA this uridine undergoes hyper-modification to 1-methyl-3-α-amino-α-carboxyl-propyl pseudouridine (m^{1}acp^{3}Ψ) (Fig. 1b)[2]. The m^{1}acp^{3}Ψ modification perturbs standard Watson-Crick base-pairing during cDNA synthesis by reverse transcriptase (RT)[23], resulting in base misincorporation and enzyme stalling, which is read-out as a consistent 'modification signature' in RNA-seq (Fig. 1c, reviewed in [23]). The increase in reference U sequence suggests that the m^{1}acp^{3}Ψ modification is incomplete or sub-stoichiometric in CRC tumors, which we term the 'hypo-m^{1}acp^{3}Ψ phenotype'. The 28S.1321.m^{1}A and 28S.4532.m^{3}U rRNA modifications also cause a 'modification signature' in RNA-seq[23]. These modifications do not decrease in CRC tumors or matched normal controls, excluding a non-specific rRNA modification effect (Fig. 1d).

The hypo-m^{1}acp^{3}Ψ phenotype is reproducible at comparable frequency (27.8-45.9%) in three additional independent patient-matched CRC cohorts (Fig. 1e,f). Analysis of 10,036 cancer patients and 712 normal controls across an additional 31 cancer-patient cohorts reveals that hypo-m^{1}acp^{3}Ψ occurs at a significant frequency across a diverse set of cancers, but is not pan-cancerous (global recurrence: 9.6% range: 0-52.8%) (Fig. 1f, S1).

To validate these findings, we designed a simple and rapid RT-PCR m^{1}acp^{3}Ψ assay for measuring 18S.1248.m^{1}acp^{3}Ψ modification. The m^{1}acp^{3}Ψ assay is reproducible and quantitative (Fig. S2). An important technical limitation is that different RT enzymes have different base misincorporation rates at 18S.1248.m^{1}acp^{3}Ψ, thus, cross-cohort or cross-experimental comparisons should be made cautiously, including in RNA-seq (Fig. S2b-d). Indeed, batch-effects on VAF are seen within TCGA cohorts, but

hypo-$m^1acp^3\Psi$ replicates across batches, further supporting that hypo-$m^1acp^3\Psi$ is occurring in tumors specifically (Fig. S1b).

We validated that the hypo-$m^1acp^3\Psi$ phenomenon also occurs in CRC cell lines assayed as a single batch and confirmed the results are not a sequencing artifact (Fig. S2e). To test if $m^1acp^3\Psi$-deficient rRNA incorporate into mature ribosomes, we isolated monosomes and polysomes and detected low $m^1acp^3\Psi$ modification levels in mono- and di-somes (Fig. S2f). As the molecular, biological, and medical significance of 18S.1248.$m^1acp^3\Psi$ unfolds, it is obvious that genotyping technologies (such as sequencing or our $m^1acp^3\Psi$ assay) and previous $m^1acp^3\Psi$ assays such as primer extension can be adapted as affordable and rapid diagnostic or prognostic assays.

18S:1248.$m^1acp^3\Psi$ is an ancient modification at the P-site core

We next investigated the evolutionary and structural characteristics of 18S.1248.$m^1acp^3\Psi$ for functional insight. The 18S:1248.U base and $m^1acp^3\Psi$ modification are absolutely conserved across *Eukarya* at a residue located in the loop of universal helix 31 (Fig. S3). *TSR3* is the aminocarboxylpropyl transferase which deposits the acp_3 at 18S.1248.U[24] and it only modifies this single rRNA position, in 100% of mature rRNA molecules[25,26].

Structurally, 18S.1248.$m^1acp^3\Psi$ is solvent-exposed at the ribosomal P-site, immediately adjacent to the codon:anti-codon interface (Fig. 2). Cryo-EM structures[27,28] and our molecular dynamics (MD) simulations implicate the m_1acp_3-modification in a direct interaction with P-site tRNA with the carboxyl-moiety forming a hydrogen bond with the universally conserved RPS16 p.R146[29] and reducing 18S rRNA flexibility at the decoding site (Fig. S4).

Start AUG-codon selection and translational initiation is a rate-limiting step in protein-synthesis and both occur at the P-site. Thus, the hypo-$m^1acp^3\Psi$ phenotype may demarcate a class of ‘onco-ribosome’ with deregulated translation. It is noteworthy that the two largest effect size RP cancer driver mutations also occur at the ribosomal P-site/tRNA interface, the RPL10 p.R98S at the peptidyl transfer site[17,18] and RPS15 C` tail mutations adjacent (<12Å) to 1248.$m^1acp^3\Psi$ (Fig. 2a) suggesting that the ribosomal P-site is a convergent multi-cancer oncogenic hot-spot.

Since the discovery of streptomycin in 1944, the ribosome has been the target of several important classes of drugs[30]. The pervasive and recurrent loss of a solvent-exposed and charged acp_3 modification at the decoding core of the small sub-unit raises the possibility that this pocket may be therapeutically

exploited with ribosome targeting antibiotics or their derivatives as a new generation of chemotherapies.

Loss of 18S.1248.m^{1}acp^{3}Ψ modification induces ribosomal protein mRNA translation

To delineate the function of 18S.1248.m^{1}acp^{3}Ψ, we generated *TSR3* knockout CRC cell lines (HCT116). Similar to yeast[24], *TSR3* is non-essential and we isolated two *TSR3* homozygous knockouts (*TSR3*[KO 1,3]), a heterozygous knockout (*TSR3*[Het 2]), as well as three wildtype control clones (WT[1-3]). Knockouts were functionally confirmed by three independent m^{1}acp^{3}Ψ assays, with *TSR3*[Het 2] showing an intermediate or hypo-m^{1}acp^{3}Ψ phenotype (Fig. 3a, S2g). Loss of the acp$_3$ modification via *TSR3*[KO] is sufficient to abolish nucleotide misincorporation during RT as measured by RNA-seq, supporting hypo-m^{1}acp^{3}Ψ tumors contain sub-stoichiometric loss of the acp$_3$-moiety.

Morphologically, the HCT116 clones were indistinguishable and showed comparable, rapid growth *in vitro* (Fig. 3b). To determine how loss of 18S.1248.m^{1}acp^{3}Ψ modification affects the transcriptome and translatome of these cells we performed RNA-seq and ribo-seq, respectively (Fig. S5).

Transcriptionally, gene set enrichment analysis (GSEA) revealed that *TSR3*[KO]/[Het] (vs. WT) cells were dominated by a proliferative tumor expression signature, characterized by elevated E2F transcription factor activity (Fig. 3c, Fig. S6, table S3). Yet, *TSR3*[KO]/[Het] cells also have a paradoxical decrease in translational efficiency of the same E2F target genes (Fig. S6).

To determine how loss of 18S.1248.m^{1}acp^{3}Ψ modification alters translation, we contrasted translational efficiency between genotypes. *TSR3*[KO]/[Het] cells showed a remarkable enrichment (vs. WT) in the translation of RPs and associated with a depletion of RP mRNA (Fig. 3d).

To validate if this RP mRNA/protein signature is present in patients, we analyzed the CPTAC-CRC cohort with tumor matched RNA-seq and proteomics data[31]. Similar to *TSR3*[KO]/[Het] cell lines, hypo-m^{1}acp^{3}Ψ CRC tumors share the same E2F oncogenic gene signature and a proteomic increase in RPs relative to normo-m^{1}acp^{3}Ψ CRC tumor controls (Fig. 3e,f).

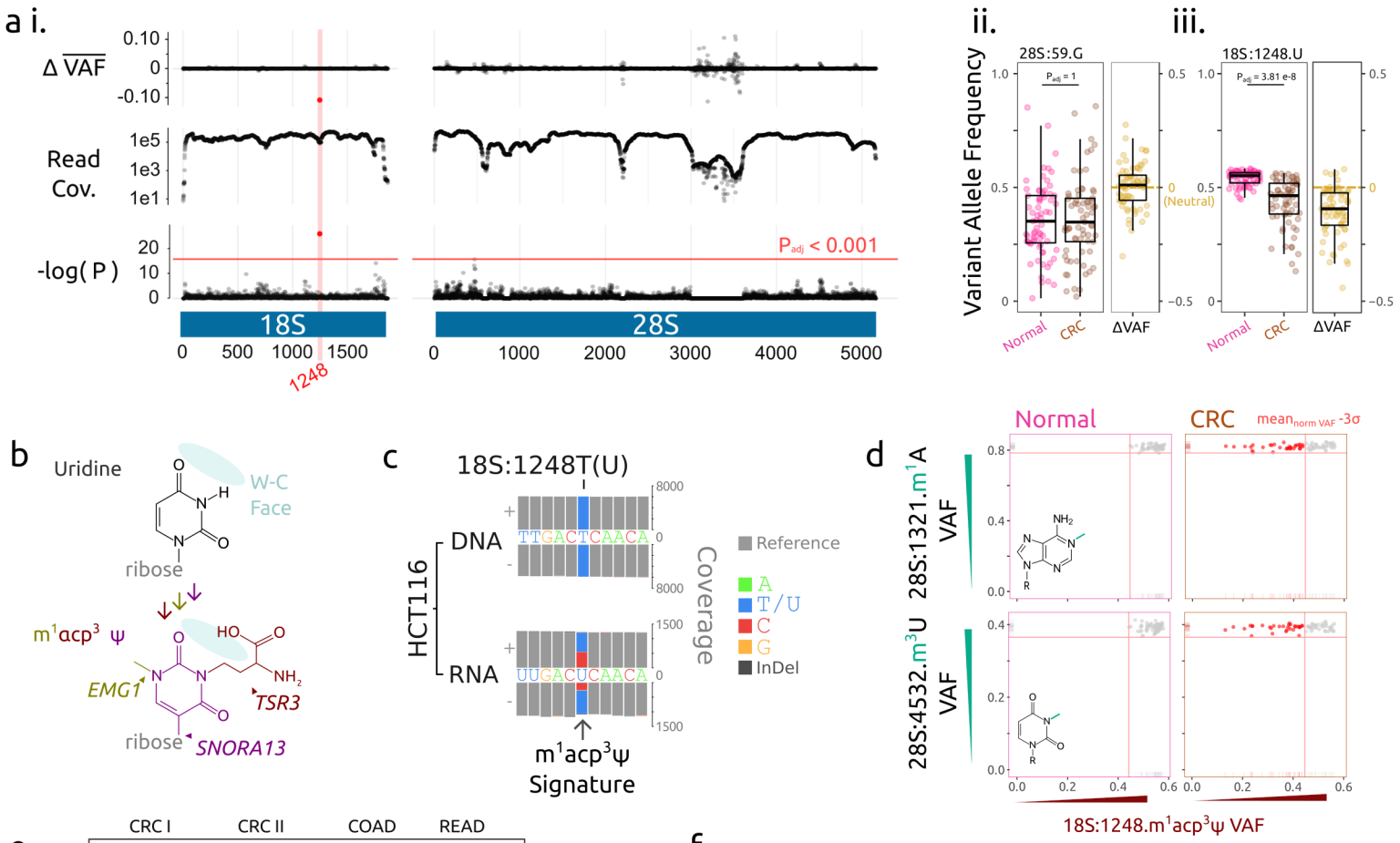
There are two hypotheses with which to interpret the hypo-m^{1}acp^{3}Ψ phenotype. The ‘oncogenic driver hypothesis’ is that m^{1}acp^{3}Ψ-deficient rRNA arise in tumorigenesis, and their dysregulated translation confers a selective advantage to the cancer, likely via high RP output. The recapitulation of the *TSR3*[KO]/[Het] multi-omic phenotype in hypo-m^{1}acp^{3}Ψ CRC patients supports a causal model. Alternatively, m^{1}acp^{3}Ψ-deficiency arises in consequence to hyper-proliferation and high ribosomal

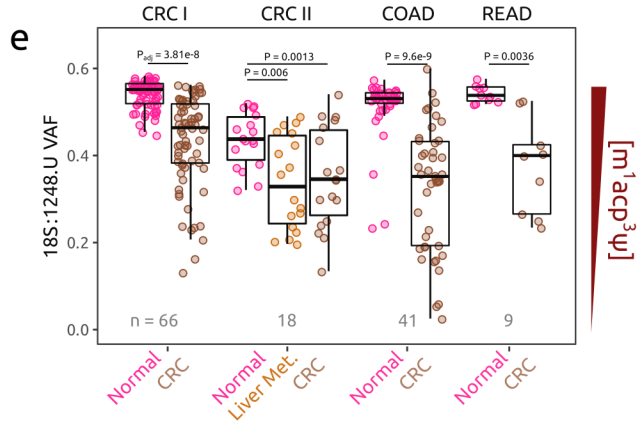
biogenesis. Rapid cellular turn-over in turn results in ‘incomplete modification’ of rRNA. Under this model the consequences of $m^1acp^3\Psi$-deficient rRNA is near-neutral or tolerably detrimental to tumor fitness. Nevertheless, hypo-$m^1acp^3\Psi$ is a highly recurrent pertubation to the ancient peptidyl-decoding core and underlies a greater cancer-translational phenomenon.

Conclusions

Ribosomes are the fulcrum in the central dogma of molecular biology. Multi-omics studies have repeatedly highlighted the discordance between mRNA and protein abundance[32,33], emphasizing the role of translational variability in physio-normal and pathological states. Several recent studies have begun to resolve the ribosome from a uniform assembly into a rich tapestry of functionally heterogeneous complexes making up distinct translational compartments in the cell[15,16,34–37]. We have discovered a pervasive and cancer-specific ‘onco-ribosome’ marked by loss of rRNA $m^1acp^3\Psi$ modification. Enticingly, the cancer-specific $m^1acp^3\Psi$-deficient ribosomes are exceptionally recurrent and can be explored as a novel chemotherapeutic class.

Figure 1


a i.
Δ VAF
Read Cov.
-log(P)
P_adj < 0.001
18S
28S
1248
ii.
iii.
28S:59.G
18S:1248.U
Variant Allele Frequency
Normal
CRC
ΔVAF
(Neutral)
b
Uridine
W-C Face
ribose
m¹acp³ ψ
EMG1
TSR3
SNORA13
c
18S:1248T(U)
HCT116
DNA
RNA
Coverage
Reference
A
T/U
C
G
InDel
m¹acp³ψ Signature
d
Normal
CRC
mean_norm VAF -3σ
28S:1321.m¹A VAF
28S:4532.m³U VAF
18S:1248.m¹acp³ψ VAF


e
CRC I
CRC II
COAD
READ
18S:1248.U VAF
P_adj = 3.81e-8
P = 0.0013
P = 0.006
P = 9.6e-9
P = 0.0036
n = 66
18
41
9
Normal
CRC
Liver Met.
[m¹acp³ψ]

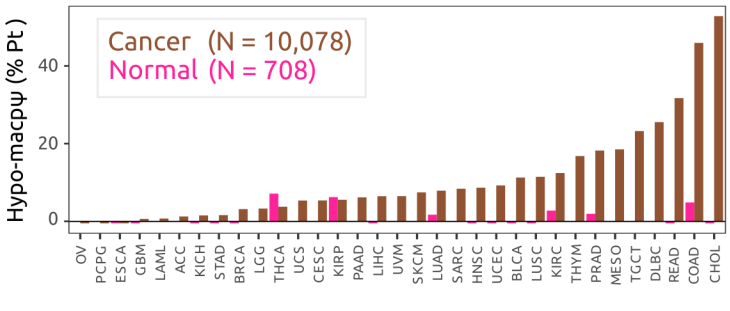

f
Hypo-macpψ (% Pt)
Cancer (N = 10,078)
Normal (N = 708)
OV
PCPG
ESCA
GBM
LAML
ACC
KICH
STAD
BRCA
LGG
THCA
UCS
CESC
KIRP
PAAD
LIHC
UVM
SKCM
LUAD
SARC
HNSC
UCEC
BLCA
LUSC
KIRC
THYM
PRAD
MESO
TGCT
DLBC
READ
COAD
CHOL

Figure 1: The hypo-$m^1acp^3\Psi$ phenotype in cancer

a i, Screen for change in the average variant allele frequency (VAF) across *18S* and *28S* ribosomal RNA (rRNA) in colorectal cancer (CRC) RNA-seq compared to patient-matched normal epithelium controls (n = 69). Read coverage and quality drops at extreme GC-content (>90%) regions of 28S, these low-coverage regions were excluded from further analysis. **ii,** The common human rRNA polymorphism 28S:r.59G>A ranges from 0.05–0.93 DNA allele frequency[10] and was expressed compatibly in the normal epithelium between variant allele frequencies (VAF) of 0.01-0.86. Neither allele is directionally selected for during cancer evolution, consistent with neutral drift (p_{adj} = 1, t = -0.44). **iii,** 18S:r.1248.U is significantly enriched (p_{adj} = 3.81e-8, t = 8.33) for the reference U allele. **b,** The 18S:r.1248.U base normally undergoes enzymatic hyper-modification to 1-methyl-3-α-amino-α-carboxyl-propyl pseudouridine ($m^1acp^3\Psi$) in three steps; *SNORA13* guided pseudouridylation; EMG1 N1-methylation; and finally 3-amino-carboxyl-propylation by TSR3. **c,** Perturbation of the Watson-Crick face of the modified base results in a distinct nucleotide misincorporation signature by reverse transcriptase in first strand cDNA synthesis, which is read out on both the sense (+) and anti-sense strand (-) of sequencing. **d,** Patient 18S:r.1248.$m^1acp^3\Psi$ hypo-modification is defined as a decrease in VAF by three standard deviations (3σ) of the matched-normal samples. Hypo-$m^1acp^3\Psi$ is not correlated with the loss of other rRNA modifications detectable by RNA-seq. **e,** The hypo-$m^1acp^3\Psi$ phenotype is replicated in three additional, independent cohorts of CRC with patient-matched adjacent normal controls, including two cohorts from The Cancer Genome Atlas (TCGA), colorectal adenocarcinoma (COAD) and rectal adenocarcinoma (READ). **f,** Hypo-modification of 18S:r.1248.$m^1acp^3\Psi$ is prevalent but not ubiquitous across the TCGA cancer cohorts (n = 10,078 patients) and largely absent from patient-matched normal controls (n = 708) (see: Fig. S1).

Figure 2

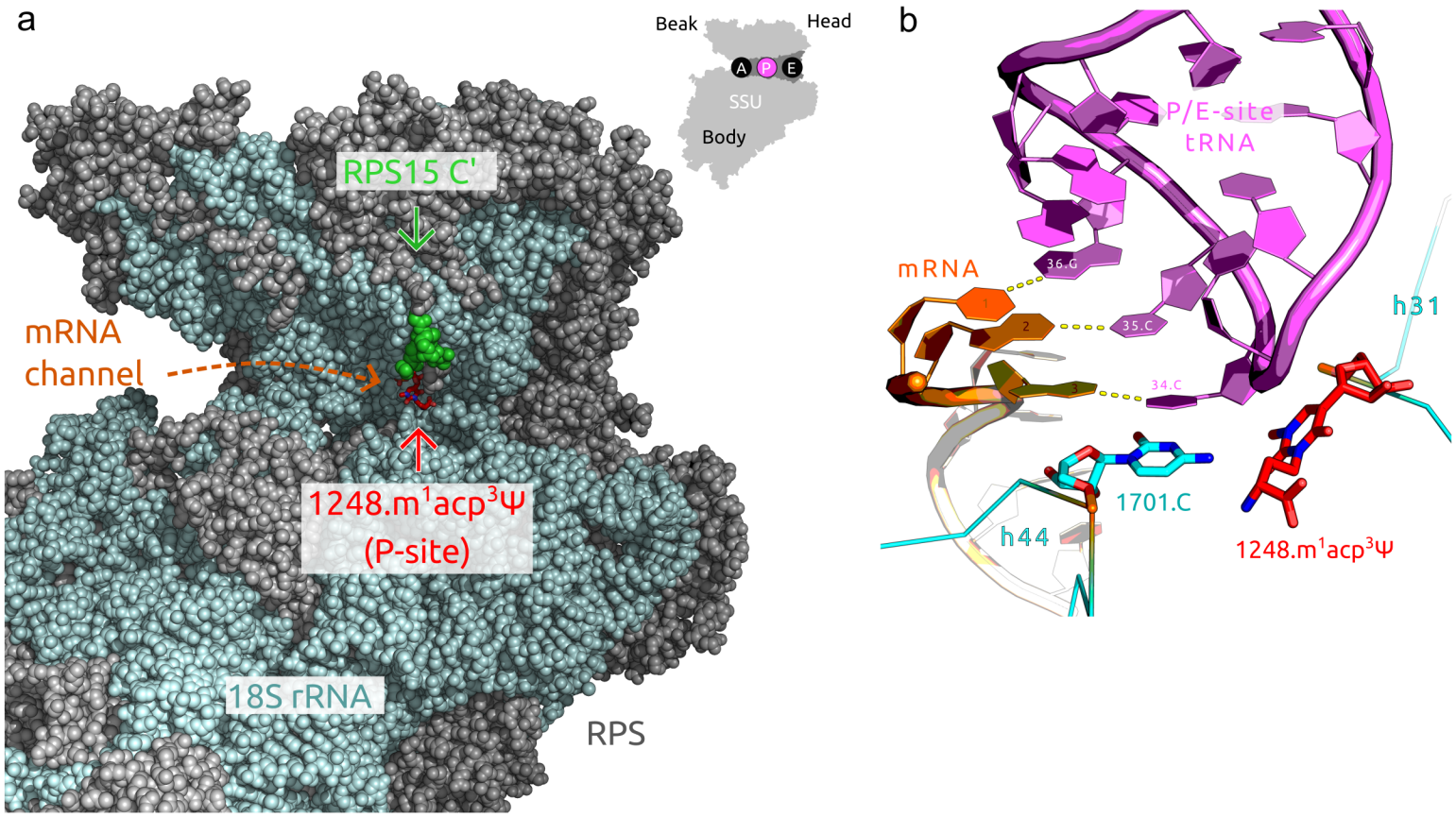

a
Beak
Head
A
P
E
SSU
Body
RPS15 C'
mRNA
channel
1248.m1acp3Ψ
(P-site)
18S rRNA
RPS
b
P/E-site
tRNA
mRNA
36.G
35.C
34.C
h31
1701.C
h44
1248.m1acp3Ψ

Figure 2: 18S:1248.m^{1}acp^{3}Ψ is located at the peptidyl decoding site

a, The mRNA channel of the human small sub-unit (SSU) cryo-EM structure with resolved base modifications (PDB: 6EK0[27]). The 18S:1248.m^{1}acp^{3}Ψ (red) nucleotide is on the loop of the universal helix 31, exposed to the mRNA channel at the center of the P-site. The CLL driver mutations in the RPS15 C` tail (green) are <12.8Å from 18S:1248.m^{1}acp^{3}Ψ. The minimal distance is likely shorter as the 10 terminal residues of RPS15 which extend into the P-site are labile and not modeled. **b,** Cryo-EM structure with a P/E-site tRNA, 18S.1248.Ψ and 18S.1701.C base stack with the ribose and base of the tRNA.34.C, respectively (PDB: 6OLE[28]). 18S.1248.m^{1}acp^{3}Ψ modification contributes to P-site decoding site stability via interaction with P-site tRNA and RPS16 (Fig. S4).

Figure 3

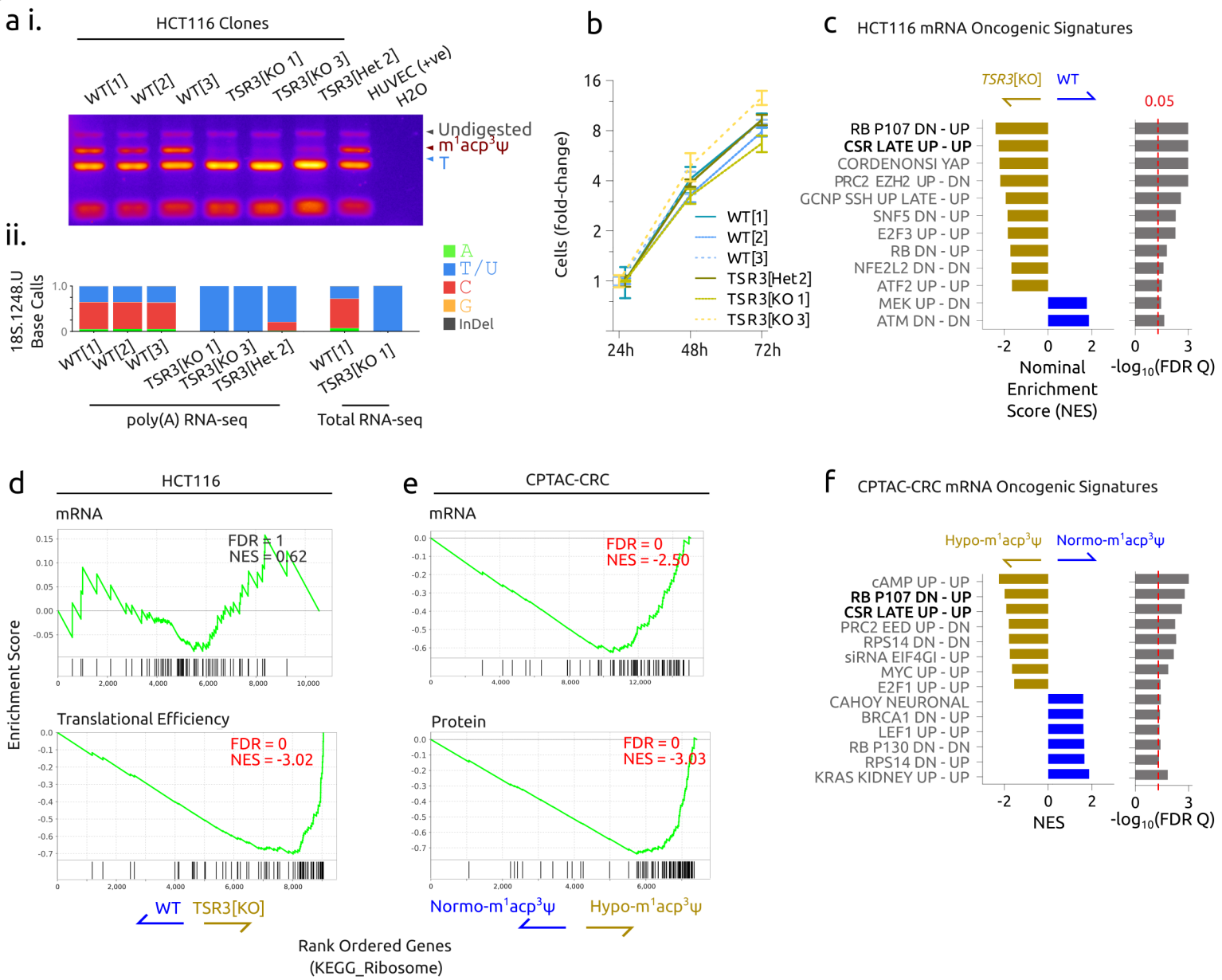

a i.
HCT116 Clones
WT[1]
WT[2]
WT[3]
TSR3[KO 1]
TSR3[KO 3]
TSR3[Het 2]
HUVEC (+ve)
H2O
Undigested
m1acp3ψ
T
ii.
18S.1248.U Base Calls
A
T/U
C
G
InDel
poly(A) RNA-seq
Total RNA-seq
b
Cells (fold-change)
24h
48h
72h
WT[1]
WT[2]
WT[3]
TSR3[Het2]
TSR3[KO 1]
TSR3[KO 3]
c
HCT116 mRNA Oncogenic Signatures
TSR3[KO]
WT
0.05
RB P107 DN - UP
CSR LATE UP - UP
CORDENONSI YAP
PRC2 EZH2 UP - DN
GCNP SSH UP LATE - UP
SNF5 DN - UP
E2F3 UP - UP
RB DN - UP
NFE2L2 DN - DN
ATF2 UP - UP
MEK UP - DN
ATM DN - DN
Nominal Enrichment Score (NES)
-log10(FDR Q)
d
HCT116
mRNA
FDR = 1
NES = 0.62
Translational Efficiency
FDR = 0
NES = -3.02
Enrichment Score
WT
TSR3[KO]
e
CPTAC-CRC
mRNA
FDR = 0
NES = -2.50
Protein
FDR = 0
NES = -3.03
Normo-m1acp3ψ
Hypo-m1acp3ψ
Rank Ordered Genes
(KEGG_Ribosome)
f
CPTAC-CRC mRNA Oncogenic Signatures
Hypo-m1acp3ψ
Normo-m1acp3ψ
cAMP UP - UP
RB P107 DN - UP
CSR LATE UP - UP
PRC2 EED UP - DN
RPS14 DN - DN
siRNA EIF4GI - UP
MYC UP - UP
E2F1 UP - UP
CAHOY NEURONAL
BRCA1 DN - UP
LEF1 UP - UP
RB P130 DN - DN
RPS14 DN - UP
KRAS KIDNEY UP - UP
NES
-log10(FDR Q)

Figure 3: The translational signature of $m^1acp^3\Psi$-deficient ribosomes

a i, Reverse transcription (RT)-PCR m^1acp^3-assay (see: methods and Fig. S2) and **ii,** RNA-seq measurement for nucleotide misincorporation at 18S:1248.$m^1acp^3\Psi$ in clones of the colorectal cancer HCT116 cell line with HUVEC as a normal positive control. **b,** The fold-change growth of HCT116 clone populations in culture, normalized to cells at 24 hours. **c,** Summary of Gene Set Enrichment Analysis (GSEA) of RNA-seq comparing HCT116 WT[1-3] versus HCT116 *TSR3*[KO 1,3 / Het 2] clones. Only significant (False Discovery Rate Q-value < 0.05) gene sets in the Oncogenic Signature collection are shown. **d,** *TSR3*[KO/Het] cells GSEA shows no change in ribosomal protein (RP) mRNA abundance, while RP translational efficiency increases. **e,** CPTAC colorectal carcinoma (CRC) tumor GSEA shows a similar increase in RP protein abundance. **f,** Summary of Oncogenic Signature GSEA comparing CPTAC-CRC tumors with normo-$m^1acp^3\Psi$ and hypo-$m^1acp^3\Psi$ modification. Gene sets common to HCT116 *TSR3*[KO]/[Het] and hypo-m^1acp^3 CPTAC-CRC patients are bolded.

References

1. Cech, T. R. The Ribosome Is a Ribozyme. *Science* **289**, 878–879 (2000).
2. Taoka, M. *et al.* The complete chemical structure of Saccharomyces cerevisiae rRNA: partial pseudouridylation of U2345 in 25S rRNA by snoRNA snR9. *Nucleic Acids Res.* **44**, 8951–8961 (2016).
3. Sloan, K. E. *et al.* Tuning the ribosome: The influence of rRNA modification on eukaryotic ribosome biogenesis and function. *RNA Biol.* **14**, 1138–1152 (2017).
4. Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
5. Elder, J. F. & Turner, B. J. Concerted evolution of repetitive DNA sequences in eukaryotes. *Q. Rev. Biol.* **70**, 297–320 (1995).
6. Leffers, H. & Andersen, A. H. The sequence of 28S ribosomal RNA varies within and between human cell lines. *Nucleic Acids Res.* **21**, 1449–1455 (1993).
7. Kuo, B. A., Gonzalez, I. L., Gillespie, D. A. & Sylvester, J. E. Human ribosomal RNA variants from a single individual and their expression in different tissues. *Nucleic Acids Res.* **24**, 4817–4824 (1996).
8. Bik, H. M., Fournier, D., Sung, W., Bergeron, R. D. & Thomas, W. K. Intra-Genomic Variation in the Ribosomal Repeats of Nematodes. *PLOS ONE* **8**, e78230 (2013).
9. Rabanal, F. A. *et al.* Unstable Inheritance of 45S rRNA Genes in Arabidopsis thaliana. *G3 GenesGenomesGenetics* **7**, 1201–1209 (2017).
10. Babaian, A. Intra- and Inter-individual genetic variation in human ribosomal RNAs. *bioRxiv* 118760 (2017) doi:10.1101/118760.
11. Kim, J.-H. *et al.* Variation in human chromosome 21 ribosomal RNA genes characterized by TAR cloning and long-read sequencing. *Nucleic Acids Res.* **46**, 6712–6725 (2018).
12. Parks, M. M. *et al.* Variant ribosomal RNA alleles are conserved and exhibit tissue-specific expression. *Sci. Adv.* **4**, eaao0665 (2018).

13. Nakhoul, H. *et al.* Ribosomopathies: Mechanisms of Disease. *Clin. Med. Insights Blood Disord.* **7**, 7–16 (2014).
14. Goudarzi, K. M. & Lindstrom, M. S. Role of ribosomal protein mutations in tumor development (Review). *Int. J. Oncol.* **48**, 1313–1324 (2016).
15. Shi, Z. & Barna, M. Translating the genome in time and space: specialized ribosomes, RNA regulons, and RNA-binding proteins. *Annu. Rev. Cell Dev. Biol.* **31**, 31–54 (2015).
16. Dinman, J. D. Pathways to Specialized Ribosomes: The Brussels Lecture. *J. Mol. Biol.* **428**, 2186–2194 (2016).
17. Girardi, T. *et al*. The T-cell leukemia-associated ribosomal RPL10 R98S mutation enhances JAK-STAT signaling. *Leukemia* **32**, 809–819 (2018).
18. Kampen, K. R. *et al.* The ribosomal RPL10 R98S mutation drives IRES-dependent BCL-2 translation in T-ALL. *Leukemia* **33**, 319–332 (2019).
19. Bretones, G. *et al*. Altered patterns of global protein synthesis and translational fidelity in RPS15-mutated chronic lymphocytic leukemia. *Blood* **132**, 2375–2388 (2018).
20. Vlachos, A., Rosenberg, P. S., Atsidaftos, E., Alter, B. P. & Lipton, J. M. Incidence of neoplasia in Diamond Blackfan anemia: a report from the Diamond Blackfan Anemia Registry. *Blood* **119**, 3815–3819 (2012).
21. Nieminen, T. T. *et al.* Germline mutation of RPS20, encoding a ribosomal protein, causes predisposition to hereditary nonpolyposis colorectal carcinoma without DNA mismatch repair deficiency. *Gastroenterology* **147**, 595–598.e5 (2014).
22. Tan, C. & Du, X. KRAS mutation testing in metastatic colorectal cancer. *World J. Gastroenterol. WJG* **18**, 5171–5180 (2012).
23. Helm, M. & Motorin, Y. Detecting RNA modifications in the epitranscriptome: predict and validate. *Nat. Rev. Genet.* **18**, 275–291 (2017).
24. Meyer, B. *et al.* Ribosome biogenesis factor Tsr3 is the aminocarboxypropyl transferase responsible for 18S rRNA hypermodification in yeast and humans. *Nucleic Acids Res.* **44**, 4304–4316 (2016).

25. Yang, J. *et al.* Mapping of Complete Set of Ribose and Base Modifications of Yeast rRNA by RP-HPLC and Mung Bean Nuclease Assay. *PLoS ONE* **11**, (2016).
26. Taoka, M. *et al.* Landscape of the complete RNA chemical modifications in the human 80S ribosome. *Nucleic Acids Res.* **46**, 9289–9298 (2018).
27. Natchiar, S. K., Myasnikov, A. G., Kratzat, H., Hazemann, I. & Klaholz, B. P. Visualization of chemical modifications in the human 80S ribosome structure. *Nature* **551**, 472–477 (2017).
28. Li, W. *et al.* Structural basis for selective stalling of human ribosome nascent chain complexes by a drug-like molecule. *Nat. Struct. Mol. Biol.* **26**, 501–509 (2019).
29. Jindal, S., Ghosh, A., Ismail, A., Singh, N. & Komar, A. A. Role of the uS9/yS16 C-terminal tail in translation initiation and elongation in Saccharomyces cerevisiae. *Nucleic Acids Res.* **47**, 806–823 (2019).
30. McCoy, L. S., Xie, Y. & Tor, Y. Antibiotics that target protein synthesis. *Wiley Interdiscip. Rev. RNA* **2**, 209–232 (2011).
31. Vasaikar, S. *et al.* Proteogenomic Analysis of Human Colon Cancer Reveals New Therapeutic Opportunities. *Cell* **177**, 1035–1049.e19 (2019).
32. Vogel, C. & Marcotte, E. M. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat. Rev. Genet.* **13**, 227–232 (2012).
33. Liu, Y., Beyer, A. & Aebersold, R. On the Dependency of Cellular Protein Levels on mRNA Abundance. *Cell* **165**, 535–550 (2016).
34. Slavov, N., Semrau, S., Airoldi, E., Budnik, B. & van Oudenaarden, A. Differential Stoichiometry among Core Ribosomal Proteins. *Cell Rep.* **13**, 865–873 (2015).
35. Xue, S. *et al.* RNA regulons in Hox 5′ UTRs confer ribosome specificity to gene regulation. *Nature* **517**, 33–38 (2015).
36. Shi, Z. *et al.* Heterogeneous Ribosomes Preferentially Translate Distinct Subpools of mRNAs Genome-wide. *Mol. Cell* **67**, 71–83.e7 (2017).
37. Fujii, K., Susanto, T. T., Saurabh, S. & Barna, M. Decoding the Function of Expansion Segments in Ribosomes. *Mol. Cell* **72**, 1013–1020.e6 (2018).

Data and code availability

Sequencing data generated in this study is available on NCBI SRA (pending accession). Electronic laboratory notebook for these experiments and analysis scripts are available at https://www.github.com/ababaian/Crown.

Acknowledgments

This work was supported by grants from the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Leukemia and Lymphoma Society of Canada to DLM. AB was supported by a National Science and Engineering Research Council (NSERC) Alexander Graham Bell Graduate Scholarship and a Roman Babicki Fellowship in Medical Research from the University of British Columbia. Sequence computation was provided by Amazon Web Services (AWS) under a Research Grant to AB. CRISPR-Cas9 reagents were provided by the Integrated DNA Technologies under the CRISPR-Challenge Prize to AB/DLM. Molecular dynamic computation was provided by Compute-Canada. DG was supported by an Alberta Innovates (Technology Futures) Graduate Student Scholarship. HJW was supported by an Alberta Innovates (Strategic Chairs Program SC60-T2) and NSERC Discovery Grant (RGPIN-2016-05199).

Data used in this publication were generated by The Cancer Genome Atlas (TCGA) Research Network (https://www.cancer.gov/tcga) and National Cancer Institute Clinical Proteomic Tumor Analysis Consortium (CPTAC).

Author contributions

A.B. discovered hypo-$m^1acp^3\Psi$ led the study and. A.B. performed DNA- and RNA-seq analysis. A.B. and K.R. performed molecular and cell culture experiments. A.B. invented the $m^1acp^3\Psi$ RT-PCR assay. S.D. performed primer extension and $m^1acp^3\Psi$ RT-PCR optimization. D.G. performed molecular dynamics and H.J.W. and A.B. helped analyze the data. I.M. prepared ribos-seq libraries, A.B. and M.M processed and analyzed the data. H.J.W., M.L., G.M., and D.L.M. provided expert advice for experiments. All authors contributed to the design of the study. A.B. prepared the manuscript and figures.

Competing interests: The authors declare no competing interests.

Correspondence and requests for materials should be addressed to A.B.

Supplementary Information is available for this paper

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

a, The 18S.1248.U variant allele frequency (VAF) from 33 TCGA patient cohorts (study abbreviations in table S1). b Batch-specific shift in the average 18S.1248.U VAF in i, TCGA-COAD and ii, TCGA-DLBC libraries. Similar to seen in the RT-PCR m^{1}acp^{3}Ψ assay (figure 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. c, The gene expression of the m^{1}acp^{3}Ψ modifying enzymes *TSR3* and *EMG1* is not decreased or lost across the TCGA cohorts.

Supplementary Figure 2: RT-PCR m^{1}acp^{3}Ψ assay and rRNA modification in cell lines

a i, The 18S.1248.m^{1}acp^{3}Ψ 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^{1}acp^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$_1$acp$_3$ modification. **e,** The m$_1$acp$_3$ RT-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^{1}acp^{3}Ψ positive control. **f i,** Polysomal fractionation and **ii,** sub-fraction m^{1}acp^{3}Ψ RT-PCR assay of the hypo-m^{1}acp^{3}Ψ cell line K562. In cells containing a mixture of +/- m^{1}acp^{3}Ψ 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^{1}acp^{3}Ψ 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.

Supplementary Figure 3: 18S.1248.m^{1}acp^{3}Ψ is absolutely conserved in *Eukarya*

a, Evolutionary conservation of 18S rRNA in the locus surrounding helix 31 in *Eukarya* and select *Archaea* and *Eubacteria* species[38,39]. Magenta arrow indicates the position homologous to

human18S:1248.U. **b-d,** The conserved secondary structure of helix 31 and its known modification sites [38,40].

Supplementary Figure 4: Ribosomal molecular dynamics and modification modeling

The $m^1acp^3\Psi$ modification stabilizes the decoding peptidyl (P-) site via a hydrogen bond with the universally conserved RPS16 (uS9) p.R146 residue. Whole ribosome molecular dynamics simulations (MD) were ran for 25ns with 18S.1248.$m^1acp^3\Psi$ or 18S.1248.U base. **a,** Root mean squared deviation (RMSD) of MD atoms show the simulations stabilize after 5ns (with ~2Å RMSD), 20ns (highlighted) was used for analysis. **b,** Root mean squared fluctuation (RMSF) of 18S rRNA **i,** adjacent to 18S.1248 and **ii,** in helix 31 shows 18S.1248. $m^1acp^3\Psi$ is less flexible (0.5Å fluctuation) relative to unmodified uridine. **c,** Minimal distance between $m^1acp^3\Psi$ (3-carboxyl oxygen) or uridine (4-oxygen) and the closest guanidinium hydrogen of RPS16 p.R146 supports that acp^3 is necessary to be within 3.5Å for hydrogen bonding and nucleotide stabilization. **d,** 20ns simulation of showing bonding between 18S.1248.$m^1acp^3\Psi$ and RPS16 p.R146 compared to the **e,** cryo-EM structure with mRNA and P-site tRNA[28]. 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.

Supplementary Figure 5: RNA-seq and ribo-seq metrics

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: figure 3). **d 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[41] 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[41]. *TSR3*[KO 1,3] libraries had less short fragments implying $m^1acp^3\Psi$-deficient ribosomes have a lower probability of being in the rotated transition state relative to WT ribosomes. **e,** Hierarchical clustering of libraries based on total translation recapitulates mRNA clustering. **f,** Differential translation of ribo-seq expressed ($RPKM_{Ribo}$ >

0.1) genes. **g,** MA-plot for total translation, highlighting the ribosome and RB/E2F gene sets. **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 clones supporting a slower global initiation rate in those two samples. **i,** 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. **j,** The $\log_2$ mRNA fold-change and $\log_2$ translation fold-change (WT / *TSR3*[KO]/[Het]), with each gene size-scaled by $RPKM_{mRNA}$. RB/E2F gene sets are highlighted, showing RP genes are more efficiently translated in *TSR3*[KO]/[Het] clones (all points below diagonal, see also figure 3). Tukey HSD test was used for testing a statistical difference between group means (* is $p < 0.05$, ** is $p < 0.001$ and *** is $p < 0.0001$)

Supplementary Figure 6: The RB/E2F transcriptional signature associated with *TSR*[KO]/[Het]

HCT116 WT[1-3] versus *TSR3*[KO]/[Het] Gene Set Enrichment analysis (GSEA) for the oncogenic signature of *RB1*, *RBL1* (p107). **a,** 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*). **b,** Rb proteins are repressors of the E2F transcription factors. Target genes of E2F were highly enriched upon *TSR3*[KO]/[Het]. **c,** The translational output (ribo-seq signal) of genes upregulated in Rb1;p107 knockout remains increased but **d,** this gene set is translated less efficiently**.**

Supplementary Table 1:

Accessions of DNA and RNA sequencing libraries used in this study. 18S.1248.U variant allele frequency (VAF) is provided for each sample.

Supplementary Table 2:

Primers and guide RNA sequences.

Supplementary Table 3:

Significant (FDR q-value < 0.05) Gene Set Enrichment Analysis results for differential transcriptomics in HCT116 WT[1-3] versus *TSR3*[KO]/[Het] with the Hallmarks, C3 promoter motif, and C6 oncogenic signatures gene sets.

Supplementary Table 4:

Significant (FDR q-value < 0.05) Gene Set Enrichment Analysis results with the C2 gene-ontology and pathways gene sets for differential transcriptomics and translational efficiency in HCT116 WT[1-3] versus *TSR3*[KO]/[Het] cell line and transcriptomics and proteomics in normo-$m^1acp^3\Psi$ versus hypo-$m^1acp^3\Psi$ CPTAC-CRC tumors.

Materials & Methods

Ribosomal sequence alignment and variant allele frequency calculations

DNA and RNA-seq libraries used in this study were prepared via poly-A selection to enrich for the ~5% of mRNA from total RNA. Since rRNA is ~80% of cellular RNA it invariably 'contaminates' RNA-seq libraries. Typically, poly-(A) RNA-seq libraries contain 3.55% (+/- 0.685, 95% CI, CRC I cohort, N = 66) of total reads aligned to rDNA. A complete list of library accessions used in this study is available in table S1[42–45].

Libraries were aligned to the *hgr1* reference rDNA sequence[10] with *bowtie2* (v. 2.3.5.1, command: *`bowtie2 --very-sensitive-local -x hgr1 -1 <read1.fq.gz> -2 <read2.fq.gz>`*)[46]. For each cohort of libraries, a genomic variant call format (GVCF) was created with *bcftools* (v. 1.9, command: *`bcftools mpileup -f hgr1.fa –max-depth 10000 -A -min-BQ 30 -b <bam.file.list>`*)[47].

GVCF files were processed in R by custom scripts to calculate variant allele frequency (VAF). VAF is defined as 1 – reference allele frequency (reference allele depth of coverage / total depth of coverage) (scripts available at https://www.github.com/ababaian/crown)

The threshold to define hypo-modification of an RNA base (including 18S.1248.$m^1acp^3\Psi$) was defined as three standard deviations below average VAF of the normal samples within the same cohort (false discovery rate = 0.00135) when available. Fixed formalin paraffin embedded (FFPE) libraries in TCGA were negative for 18S.1248.$m^1acp^3\Psi$, 28S.1321.m^1A and 28S.4532.m^3U modification signatures and excluded from further analysis. In the CPTAC-CRC cohort (normal RNA-seq is unavailable), hypo-$m^1acp^3\Psi$ and normo-$m^1acp^3\Psi$ was defined by the lower (<25%) and upper (>75%) quantiles of samples within a batch.

Transcriptome and translatome alignment, assembly and differential expression

RNA-seq reads were aligned to *hg38* (*GRCh38)* reference genome with *tophat2* (v.2.0.14)[48]. Individual transcriptome assemblies for HCT116 [WT 1-3], [KO 1,2] and [Het 2] libraries were generated with *stringtie* (v 2.0)[49], and then all merged together with the human *gencode basic* gene annotation (v. 31)[50] ultimately yielding the *hct116_gencode.v31* reference gene set.

To generate a single-copy reference transcriptome for ribo-seq analysis of HCT116, isoform-specific quantification of gene expression was performed on the *hct116_gencode.v31* gene set with *`stringtie -G hct116_gencode.gtf`*. For each gene with non-zero expression (>10 unique reads), the one highest

expression isoform (average expression from each clone) was chosen as the reference transcript for that gene.

For ribo-seq alignment, after read adapter trimming and alignment to *hgr1* as above, unmapped reads were aligned against a containment file containing human tRNA, mtDNA, snoRNA, snRNA and miRNA sequences. Reads remaining unmapped were then aligned to *hg38* and the *hct116_transcriptome* with *STAR* aligner (v. 2.5.2b, command: `*STAR --genomeDir hg38 --readFilesIn <input.fq> --sjdbFileChrStartEnd hg38/sjdbList.out.tab --outFilterMultimapNmax 10 --outFilterMismatchNmax 5 --outFilterMatchNmin 15 --alignSJoverhangMin 5 --seedSearchStartLmax 20 --outSJfilterOverhangMin 30 8 8 8 --quantMode TranscriptomeSAM*`)[51]. Transcriptome aligned Ribo-seq data was analyzed in *R* (v. 3.5.1) using the *riboWaltz* package (v. 1.1.0)[52].

Gene-level expression and total translation was quantified with the *DEseq2*[53] R package using *hg38* aligned bam files and the *hct116_gencode.v31* reference gene set. Translational efficiency was calculated per genotype as log2(Ribo-seq $Gene_{RPKM}$ / RNA-seq $Gene_{RPKM}$).

Gene expression and translation differences were calculated by Gene Set Enrichment Analysis (*GSEA*, v.4.0.0)[54] with `*-permute gene_set -nperm 5000*` and standard parameters. Transcriptomic GSEA was performed using the MSigDB[55] (v 7.0): hallmark, C2 pathways, C3 motif search, and C6 oncogenic signatures gene sets. Translatomic and proteomic GSEA was performed with C5 Gene Ontology (GO) gene set.

All bioinformatic analyses were scripted for reproducibility and are available at https://www.github.com/ababaian/crown.

HCT116 cell culture and TSR3 knockout

The colorectal carcinoma cell line HCT116 (CCL-247, ATCC, Manassas, VA) was cultured in DMEM media (#36250, STEMCELL Technologies, Vancouver, Canada) supplemented with 10% fetal bovine serum (F1051, Invitrogen, Waltham, MA).

To generate *TSR3* knockouts, 10^5 HCT116 cells were transfected with 10 nmol of one of three *TSR3* targeting Alt-R CRISPR-Cas9 ribonucleoproteins or non-targeting controls (table S2) by manufacturer's protocol (1081059, Integrated DNA Technologies (IDT), Coralville, IA). After 24 hours, single cells from each treatment group were isolated by limiting dilution and confirmed to be 1 cell /well by microscopy. Single cell clones were expanded to $5x10^5$ cells at which point half the culture was frozen

(culture media + 10% DMSO) and half were processed for RNA. *TSR3* knockouts were genotyped by RNA-seq and functional knockout was confirmed by three independent $m^1acp^3\Psi$ assays (figure 3, S2).

Cell lines and clonal isolates were tested to be free of mycoplasma contamination by DAPI staining and microscopy and with LookOut Mycoplamsa Detection Kit (MP0035, MilliporeSigma, Burlington, MA) by manufacturer's protocol.

RNA isolation and RNA-seq

Cells for RNA extraction were lysed directly in TRIzol reagent (15596-018, Invitrogen), spun 5 min at 12,000 x g to pellet fat and nuclear DNA and then frozen at -80°C. RNA extraction was carried out by manufacturer's protocol. RNA quality was assessed via 2% denaturing RNA agarose gel electrophoresis (heat treated, 95°C for 5 minutes in 1.5x formamide loading buffer[56]) and concentration/purity assessed by spectrophotometer (NanoDrop 2000, ThermoFisher, Waltham, MA). RNA quality for RNA-seq library preparation had a >9.9 RIN score measured by Bioanalyzer 2100 (Agilent, Santa Clara, CA).

RNA-seq library preparation and sequencing was performed by the BC Cancer Genome Sciences Centre, Vancouver, Canada. Briefly, 75-bp stranded and paired-end poly-(A) RNA-seq libraries were prepared with NEBNext poly(A) mRNA magnetic isolation module (E7490L, New England BioLabs (NEB), Ipswich, MA), Maxima H minus First Strand cDNA synthesis kit (K1652, Thermo-Fisher), and NEBNext Ultra II directional RNA second strand synthesis (E7771, NEB). The total RNA-seq libraries were prepared in parallel but without poly-(A) selection and only 2x PCR cycles (for adapter ligation). Libraries were sequenced on a HiSeq 2500 (Illumina, San Diego, CA).

Assays for 18S.1248.$m^1acp^3\Psi$ modification

Primer extension was performed with 1 µg of total RNA, incubated with 2 pmol of *PE_1248_BLOCK* (IDT) primer and 2U of RNase H (18021-014, Invitrogen) or mock enzyme treatment at 37°C for 20 min followed by heat inactivation at 65°C for 10 min. SuperScript III reverse transcriptase (18080044, lot #2042663, Invitrogen) and the flourophore labeled *PE_1248_FAM* primer were added for primer annealing and RT (1h at 50°C) as described by Schuster and Bertram[57]. Labeled cDNAs were re-suspended in 1.5x formamide loading buffer and heated to 95°C for 3 min to eliminate secondary structures[56]. Samples were separated on a 2% agarose gel at 114 V for 3h at 4°C or on a 12.5% polyacrylamide gel at 45 mA for 2.5h in 1x TBE. After migration, the gel was visualized with the

Typhoon FLA 9500 laser scanner (FAM filter, 50 µm pixel and 450 V unless otherwise noted, GE Healthcare, Chicago, IL).

The RT-PCR 1248.m^{1}acp^{3}Ψ assay was performed with 1 ug of DNase treated (AM1907, lot #00733051, Invitrogen) RNA after total RNA quality was assessed by denaturing agarose gel electrophoresis[56]. RT reaction was carried out with SuperScript III (Invitrogen), SuperScript IV (18090010, lot #00721480, Invitrogen), UltraScript 2.0 (PB30.31-10, lot #PB130614-01-5, PCR Biosystems, Wayne PA) and WarmStart RTx (M0380L, lot #0061705, NEB) by each manufacturer's protocol with minor modifications. RT reactions were carried out with a random hexamer primer only, and not poly(T) oligos. cDNAs were diluted five-fold and used as template for PCR (30 cycles: 94°C for 30 s, 55°C for 30 s, 72°C for 30 s) with *macp_F1* and *macp_R1* primers (table S2). Amplicons were digested with HinFI (R0155S, New England BioLabs) (25°C for 5 sec, 37°C for 90 min, 80°C for 20 min). Samples were separated on a 2.25% agarose gel in 1x TBE at 200 V for 45 min at 4°C. After migration, the gel was post-stained in 1x GelRed (41003, Biotium, Fremont CA) for 30 min. Gels were visualized by UV transillumination, captured in gray scale with a digital camera and pseudo-colored in ImageJ[58] (v 1.52h, Lookup table > Fire) which retains the original pixel intensity values but highlights band-intensity visualization.

Ribosome foot-printing

Ribosome foot printing (ribo-seq) was performed as previously described[59] with minor modifications. For cell harvesting, the culture medium was aspirated, cells were washed twice with ice-cold PBS supplied with 100 µg/ml cycloheximide and plates were flash-frozen in liquid nitrogen. For cell lysis, the plates were placed on wet ice and 400 µl of mammalian polysome buffer (MPB) [20 mM Tris-HCl, pH 7.4, 150 mM NaCl, 5 mM $MgCl_2$, with 1 mM DTT and 100 µg/ml cycloheximide, 1% (vol/vol) Triton X-100, 25 U/ml Turbo DNase (AM2238, Invitrogen) was dripped onto the plates. Cells were scraped, the lysate was collected to fresh 1.5 ml tube, passed ten times through a 26-gauge needle, cleared by centrifugation at 20,000 x g for 10 min, flash-frozen in liquid nitrogen and stored at −80 °C until further use. For isolation of ribosome-protected RNA fragments, 240 µl of the lysate was digested with 6 µl of RNase I (AM2294, 100 U/µl, Invitrogen) at room temperature with rotation. After 45 min 8 µl of SUPERase-In (20 U/µl, AM2694, Invitrogen) was added to reaction and passed through MicroSpin S-400 HR columns (27-5140-01, GE Healthcare) equilibrated with mammalian polysome buffer. RNA was extracted from the flow-through using Trizol LS (10296-010, Invitrogen) followed by

depletion of ribosomal RNA fragments with the RiboZero Kit (MRZH11124, Illumina). Ribosome-protected RNA fragments were loaded onto denaturing 17% urea-PAGE gel (EC-829, National Diagnostics) and gel area ranging from 27 nt to 30 nt, defined by corresponding RNA markers, was cut out. Purified RNA fragments were subjected to library generation using 3′ adapter 4N-RA3, 5′ adapter OR5-4N, RT primer RTP and PCR primers RP1 (forward primer) and RPI1-15 (reverse primers, containing barcodes). Libraries were sequenced on a HiSeq 4000 device (Illumina).

Polysome Fractionation

Polysome fractionation was performed as previously described[60], with minor modifications. Media was removed from 100 mm dish with ~10^7 cells and washed with ice-cold ddH_2O containing 100 µM CHX. All subsequent steps were performed chilled at 4°C or on ice. After ddH_2O aspiration, cells were incubated for 30 min in 450 µL of hypotonic lysis buffer [0.1x polysome base buffer (PBB), 150 mM KCl, 20 mM Tris-HCl pH 7.4, 15 mM $MgCl_2$ in ddH_2O; with 1% Triton-X 100 and 1x protease inhibitor (4693132001, MilliporeSigma). After confirming >95% free nuclei with a hemocytometer, nuclei were pelleted by centrifugation at 1,800 x g for 5 minutes. Cytoplasmic fraction was separated from mitochondria by centrifugation at 10,000 x g for 5 minutes. 300 µL cytoplasmic lysate was layered atop at 7-45% sucrose gradient (Gradient Master, BioComp, Fredericton, Canada) in 1x PBB. Gradients were ultra-centrifuged at 221,600 x g for 2 hours at 4°C (SW-41Ti rotor, 331362, Beckman Coulter, Brea, CA). Gradients were fractionated (Piston Fractionator, BioComp) into 20 x 300 µL fractions with in-line UV-scanning at 254 nm. Fractions were immediately frozen at -20°C for subsequent RNA extraction.

Ribosomal molecular dynamics simulations

All molecular dynamics (MD) simulations were performed as described in Girodat et al. 2019[61]. In brief, 80S ribosome models were derived from available human Cryo-EM structures with a resolved 18S.1248.$m^1acp^3\Psi$ (PDB: 6EKO for E-site tRNA and 6OLE for A/P and P/E tRNA) [27,28].

For simulations lacking $m^1acp^3\Psi$ modifications, the base was converted to uracil. Each system was protonated with the *psfgen* package in VMD 1.9.3, and only e-nitrogen for histidine were protonated[62]. Each system was solvated with a 10Å TIP3P water box with a concentration of 7mM $MgCl_2$ and 100mM KCl using the *solvate* and *autoionize* packages, respectively[62]. All minimizations and MD simulations were performed with *NAMD* 2.1.2 using *CHARMM 36* standard parameters[63–65] and modified nucleic acid parameters from Xu et al. 2016[66].

Each system underwent a steepest descent minimization of water for 10,000 steps then water and ions for 100,000 steps twice followed by minimization of nucleic acid and protein for 50,000 steps and finally the whole system for 100,000 steps. After minimization all systems were equilibrated to 300 and 350 K for 150 ps. Coordinates of the 350 K equilibration in conjunction with velocities from the 300 K equilibration were used as initial parameters for the MD simulation. Each system was simulated for ~20ns. Energy contributions of 18S.1248.m^{1}acp^{3}Ψ or 18S.1248.U were determined with the *NAMD Energy* package[65].

Statistics

Statistical analysis was performed in *R* (v 3.5.1). Differences in variant allele frequency (VAF) between tumor and normal patient samples was two-tailed, paired Student's T-test with degrees of freedom one less than reported *n*. Bonferonni multiple-testing correction was applied when screening for changes across 18S and 28S nucleotides. Error bars on boxplots are quantiles. Differential gene expression and translation was tested with *DEseq2*[53] with Benjamini-Hochberg multiple testing correction at an alpha of 0.05. Multi-group comparisons between HCT116 WT[1],[2],[3] and *TSR3*[KO 1],[KO 3],[Het2] ribo-seq were performed with one-way ANNOVA, followed by Tukey's Honestly Significant Differencce (HSD) test if indicated.

505 **Extended References**

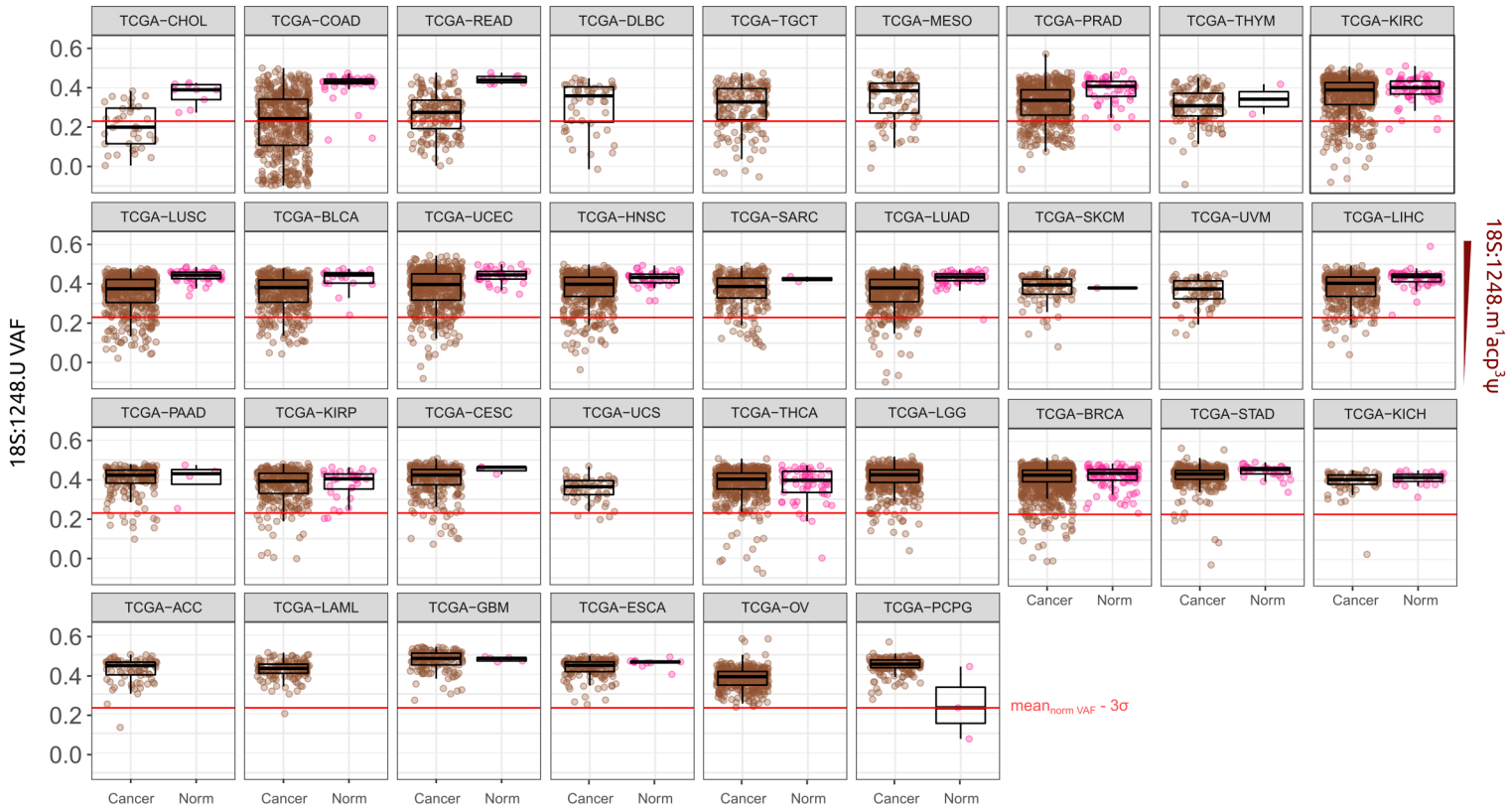
38. Petrov, A. S. *et al.* Evolution of the ribosome at atomic resolution. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 10251–10256 (2014).
39. Bernier, C. R. *et al.* RiboVision suite for visualization and analysis of ribosomes. *Faraday Discuss.* **169**, 195–207 (2014).
40. Machnicka, M. A. *et al.* MODOMICS: a database of RNA modification pathways--2013 update. *Nucleic Acids Res.* **41**, D262-267 (2013).
41. Lareau, L. F., Hite, D. H., Hogan, G. J. & Brown, P. O. Distinct stages of the translation elongation cycle revealed by sequencing ribosome-protected mRNA fragments. *eLife* **3**, e01257 (2014).
42. Seshagiri, S. *et al.* Recurrent R-spondin fusions in colon cancer. *Nature* **488**, 660–664 (2012).
43. Lim, B. *et al.* Genome-wide mutation profiles of colorectal tumors and associated liver metastases at the exome and transcriptome levels. *Oncotarget* **6**, 22179–22190 (2015).
44. Cancer Genome Atlas Research Network *et al.* The Cancer Genome Atlas Pan-Cancer analysis project. *Nat. Genet.* **45**, 1113–1120 (2013).
45. Ghandi, M. *et al.* Next-generation characterization of the Cancer Cell Line Encyclopedia. *Nature* **569**, 503–508 (2019).
46. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
47. Danecek, P. *et al.* The variant call format and VCFtools. *Bioinforma. Oxf. Engl.* **27**, 2156–2158 (2011).
48. Kim, D. *et al.* TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36 (2013).
49. Pertea, M. *et al.* StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **33**, 290–295 (2015).
50. Frankish, A. *et al.* GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res.* **47**, D766–D773 (2019).
51. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

52. Lauria, F. *et al.* riboWaltz: Optimization of ribosome P-site positioning in ribosome profiling data. *PLOS Comput. Biol.* **14**, e1006169 (2018).
53. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
54. Subramanian, A. *et al.* Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci.* **102**, 15545–15550 (2005).
55. Liberzon, A. *et al.* The Molecular Signatures Database Hallmark Gene Set Collection. *Cell Syst.* **1**, 417–425 (2015).
56. Masek, T., Vopalensky, V., Suchomelova, P. & Pospisek, M. Denaturing RNA electrophoresis in TAE agarose gels. *Anal. Biochem.* **336**, 46–50 (2005).
57. Schuster, C. F. & Bertram, R. Fluorescence based primer extension technique to determine transcriptional starting points and cleavage sites of RNases in vivo. *J. Vis. Exp. JoVE* e52134 (2014) doi:10.3791/52134.
58. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **9**, 671–675 (2012).
59. Ingolia, N. T., Brar, G. A., Rouskin, S., McGeachy, A. M. & Weissman, J. S. The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments. *Nat. Protoc.* **7**, 1534–1550 (2012).
60. Floor, S. N. & Doudna, J. A. Tunable protein synthesis by transcript isoforms in human cells. *eLife* **5**, e10921 (2016).
61. Girodat, D., Mercier, E., Gzyl, K. E. & Wieden, H.-J. Elongation Factor Tu's Nucleotide Binding Is Governed by a Thermodynamic Landscape Unique among Bacterial Translation Factors. *J. Am. Chem. Soc.* **141**, 10236–10246 (2019).
62. Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38, 27–28 (1996).

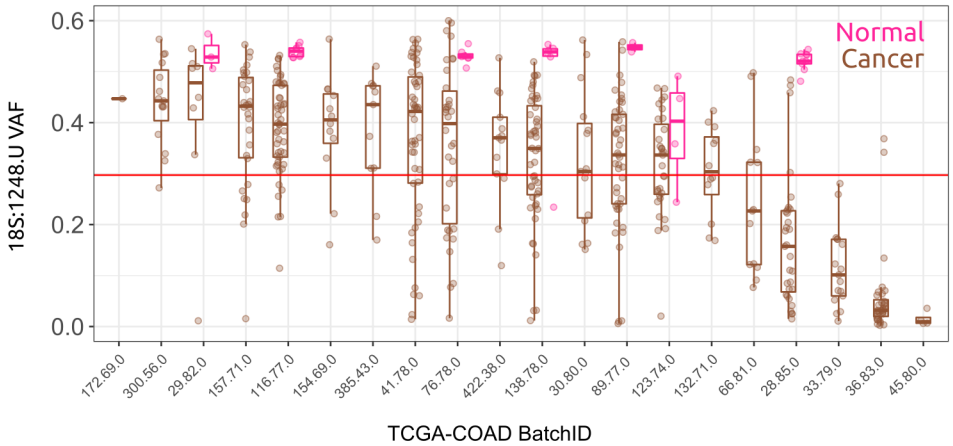
63. Best, R. B. *et al.* Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone φ, ψ and side-chain χ(1) and χ(2) dihedral angles. *J. Chem. Theory Comput.* **8**, 3257–3273 (2012).
64. Denning, E. J., Priyakumar, U. D., Nilsson, L. & Mackerell, A. D. Impact of 2'-hydroxyl sampling on the conformational properties of RNA: update of the CHARMM all-atom additive force field for RNA. *J. Comput. Chem.* **32**, 1929–1943 (2011).
65. Phillips, J. C. *et al.* Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26**, 1781–1802 (2005).
66. Xu, Y., Vanommeslaeghe, K., Aleksandrov, A., MacKerell, A. D. & Nilsson, L. Additive CHARMM force field for naturally occurring modified ribonucleotides. *J. Comput. Chem.* **37**, 896–912 (2016).

Figure S1

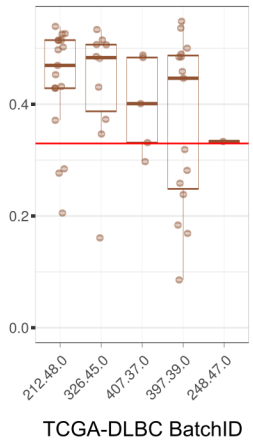
a



b i.



ii.



c

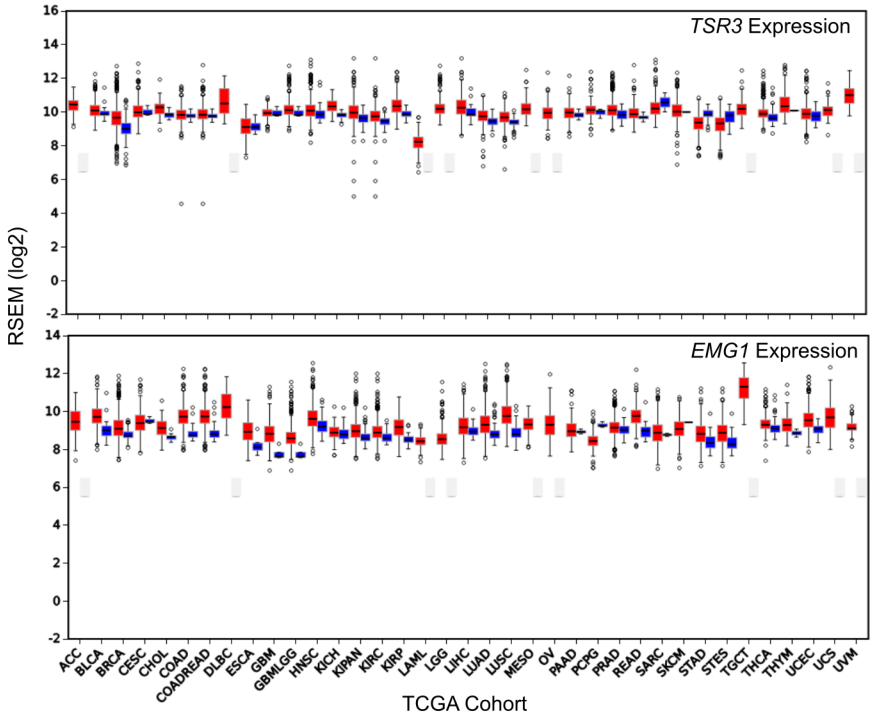


Figure S2

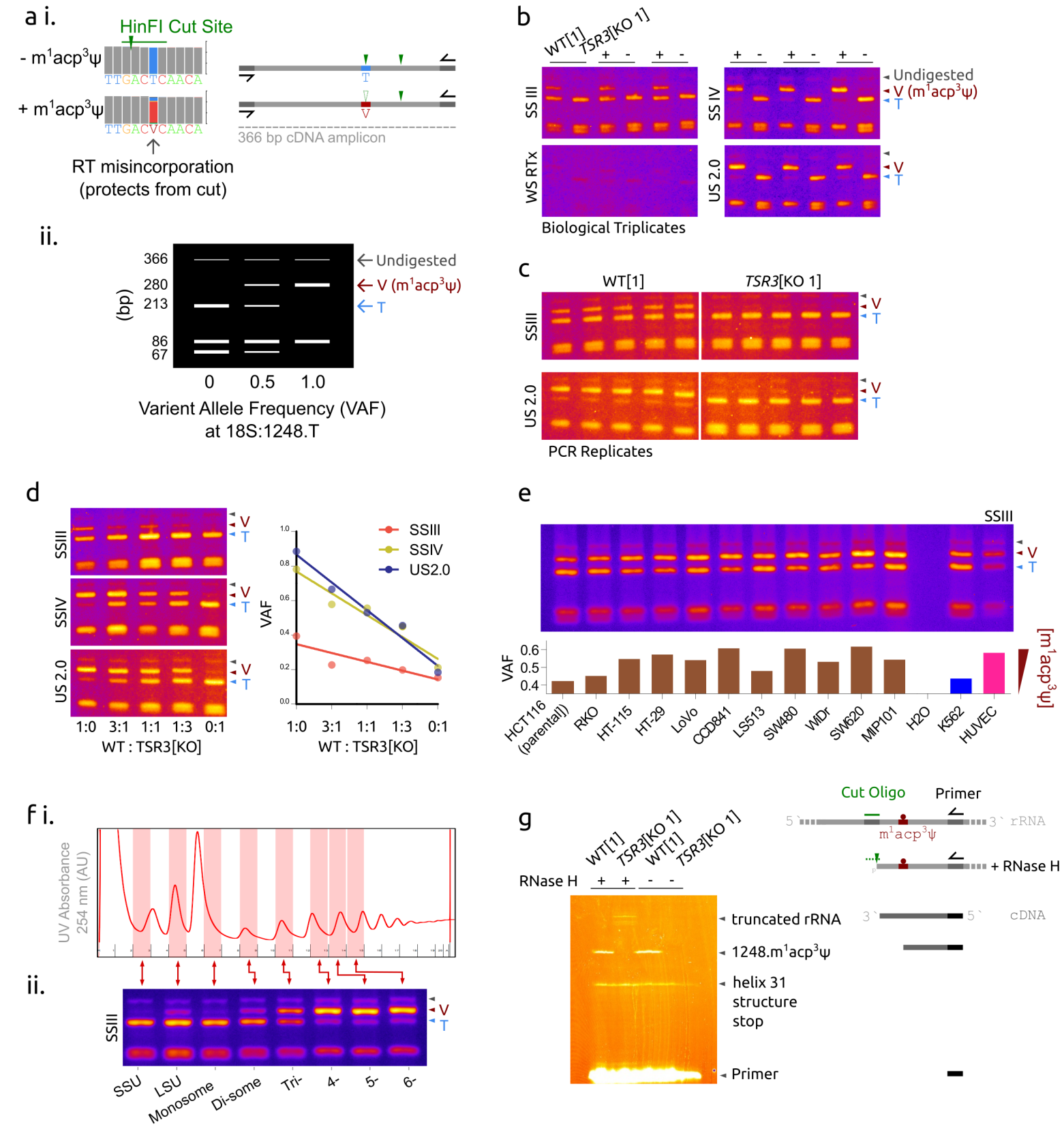
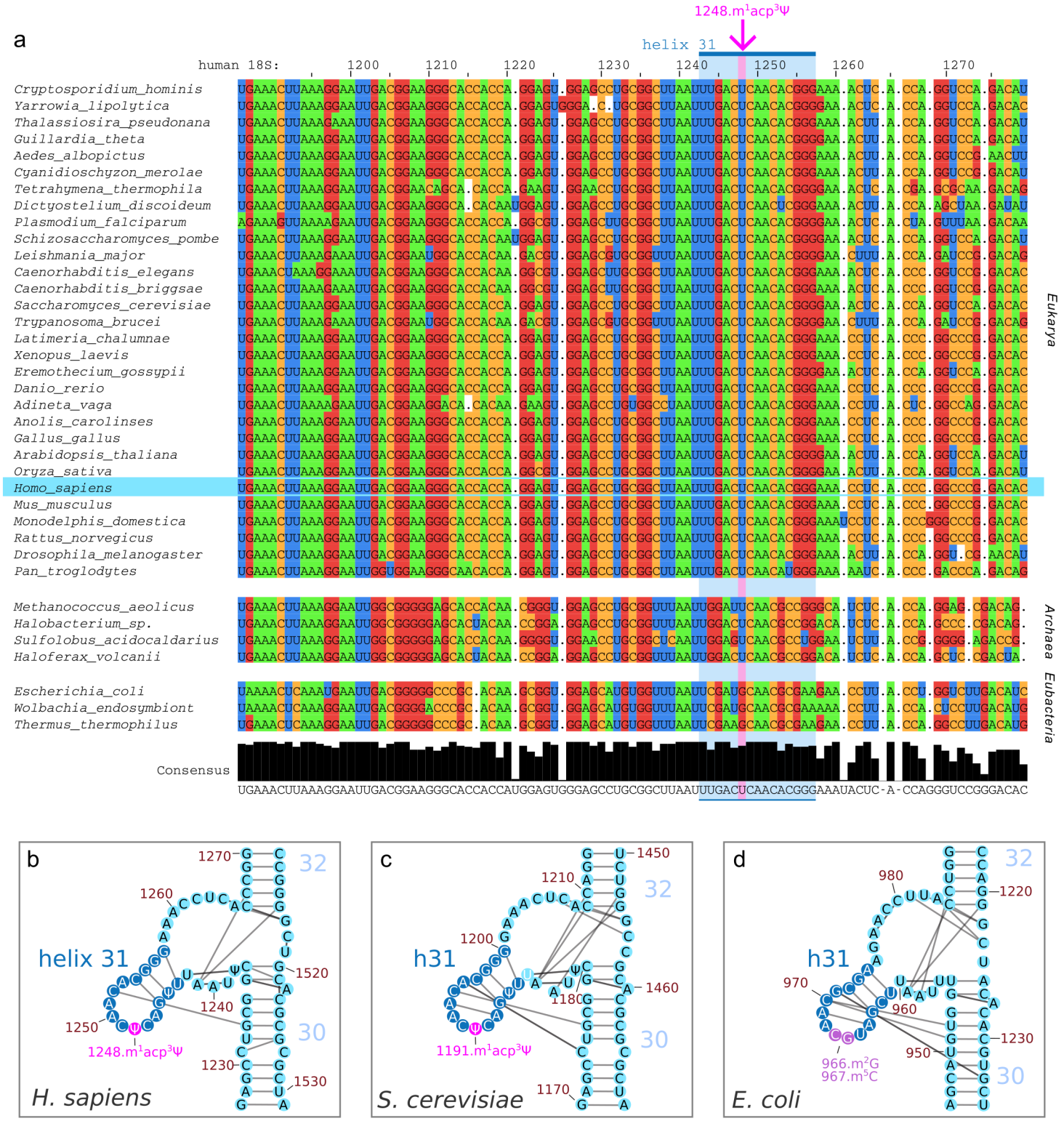
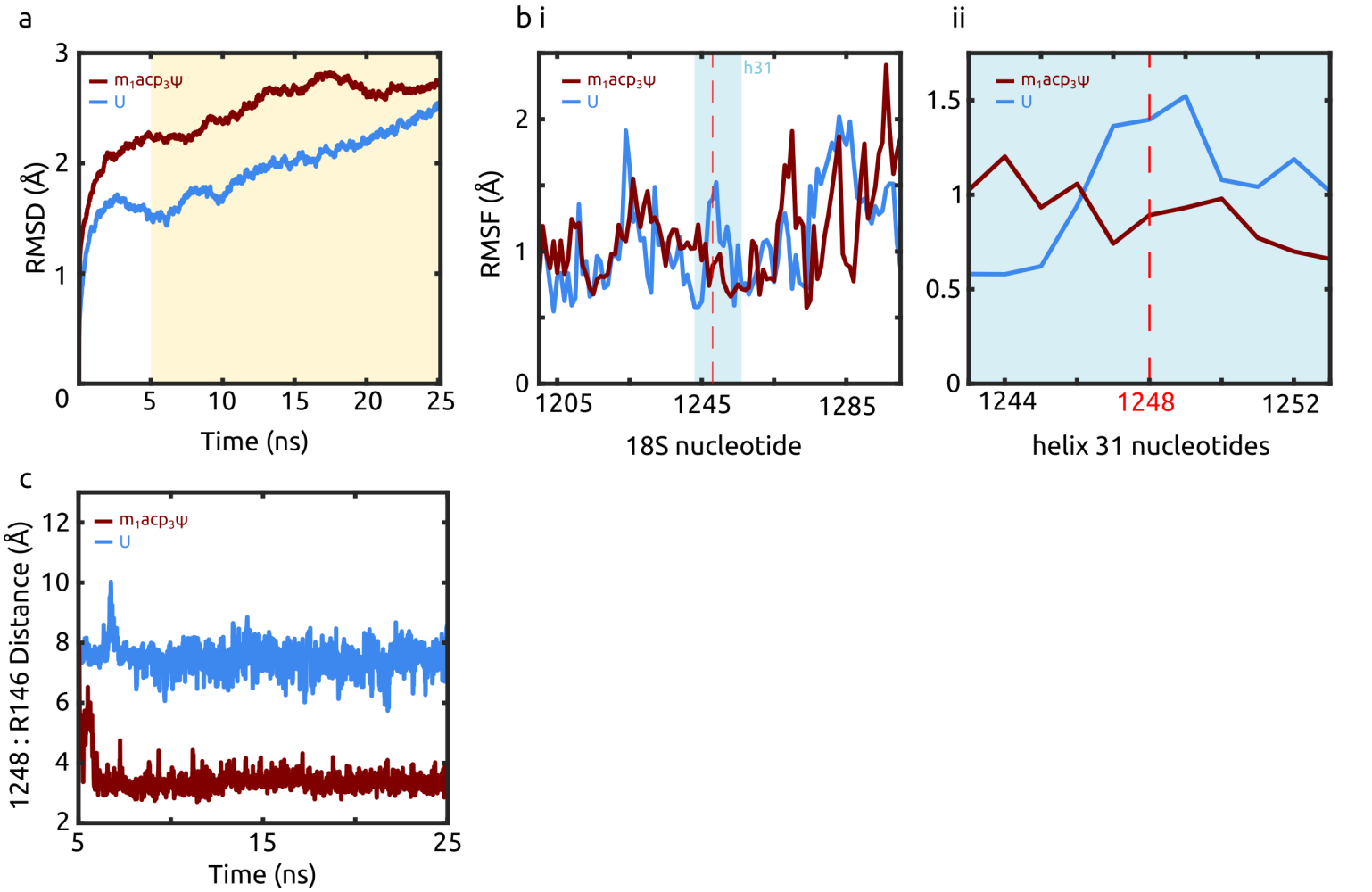

a i.
HinFI Cut Site
- m1acp3ψ
TTGACTCAACA
+ m1acp3ψ
TTGACVCAACA
T
V
366 bp cDNA amplicon
RT misincorporation
(protects from cut)
ii.
366
280
213
86
67
(bp)
Undigested
V (m1acp3ψ)
T
0
0.5
1.0
Varient Allele Frequency (VAF)
at 18S:1248.T
b
WT[1]
TSR3[KO 1]
+ - + -
+ - + - + -
SS III
SS IV
WS RTx
US 2.0
Undigested
V (m1acp3ψ)
T
V
T
Biological Triplicates
c
WT[1]
TSR3[KO 1]
SSIII
US 2.0
V
T
PCR Replicates
d
SSIII
SSIV
US 2.0
V
T
1:0 3:1 1:1 1:3 0:1
WT : TSR3[KO]
VAF
SSIII
SSIV
US2.0
1:0 3:1 1:1 1:3 0:1
WT : TSR3[KO]
e
SSIII
V
T
VAF
0.6
0.5
0.4
[m1acp3ψ]
HCT116 (parental))
RKO
HT-115
HT-29
LoVo
CCD841
LS513
SW480
WiDr
SW620
MIP101
H2O
K562
HUVEC
f i.
UV Absorbance 254 nm (AU)
ii.
SSIII
V
T
SSU
LSU
Monosome
Di-some
Tri-
4-
5-
6-
g
Cut Oligo
Primer
5` 3` rRNA
m1acp3ψ
+ RNase H
3` 5` cDNA
WT[1]
TSR3[KO 1]
WT[1]
TSR3[KO 1]
RNase H + + - -
truncated rRNA
1248.m1acp3ψ
helix 31
structure
stop
Primer

Figure S3



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
Consensus
Eukarya
Archaea
Eubacteria
b
helix 31
1248.m1acp3Ψ
H. sapiens
c
h31
1191.m1acp3Ψ
S. cerevisiae
d
h31
966.m2G
967.m5C
E. coli

Figure S4


a
RMSD (Å)
Time (ns)
m1acp3ψ
U
b i
RMSF (Å)
h31
18S nucleotide
ii
helix 31 nucleotides
1244
1248
1252
c
1248 : R146 Distance (Å)
Time (ns)

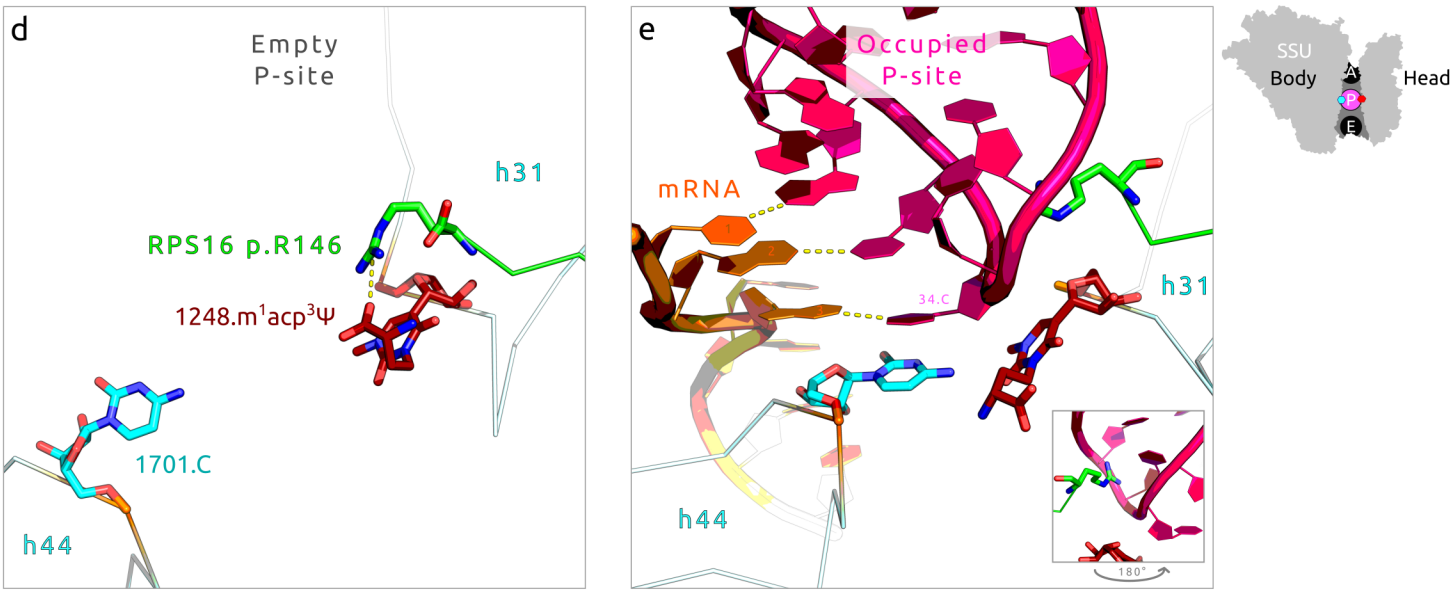

d
Empty
P-site
h31
RPS16 p.R146
1248.m1acp3ψ
1701.C
h44
e
Occupied
P-site
mRNA
34.C
h31
h44
180°
SSU
Body
Head
A
P
E

Figure S5

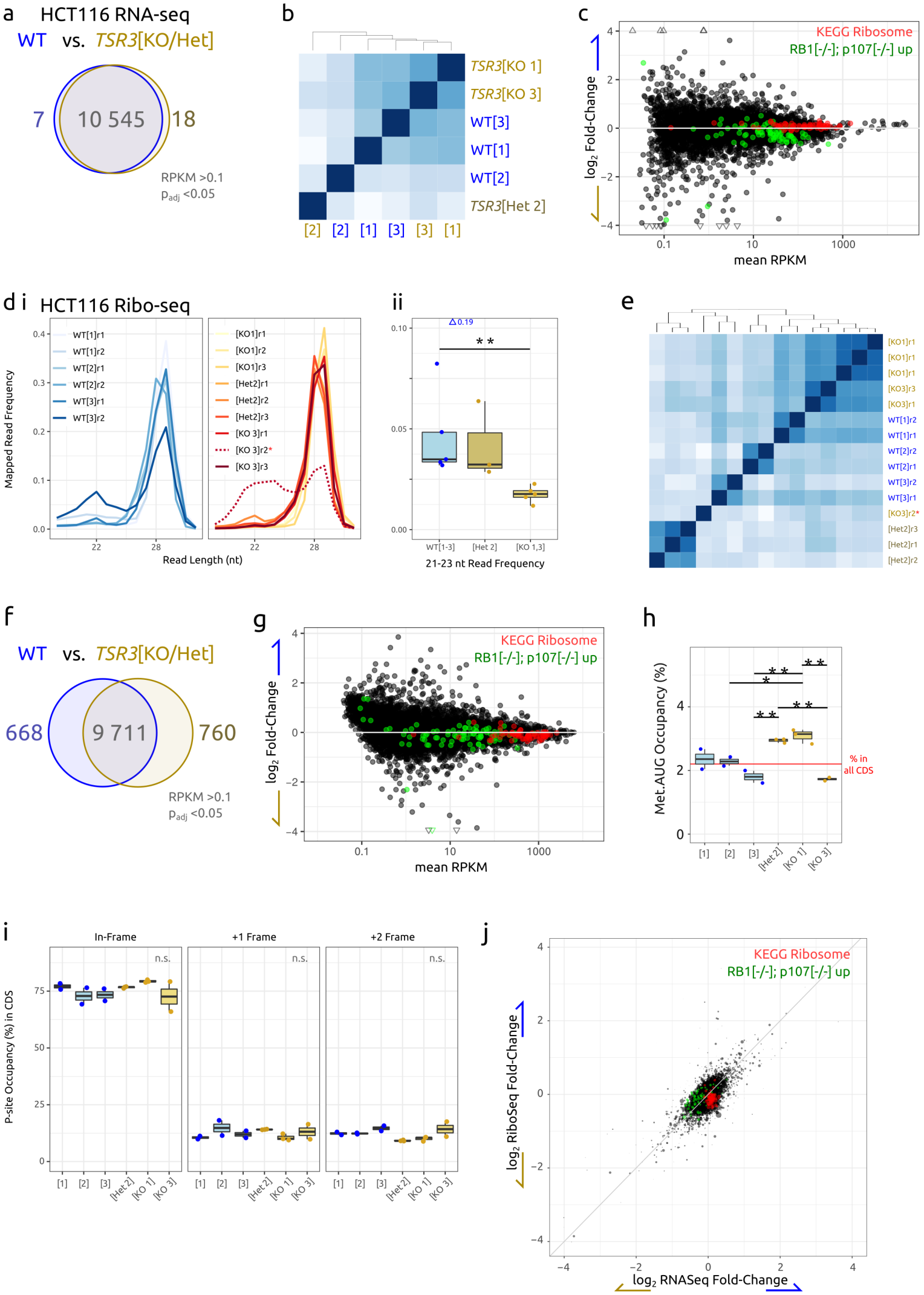
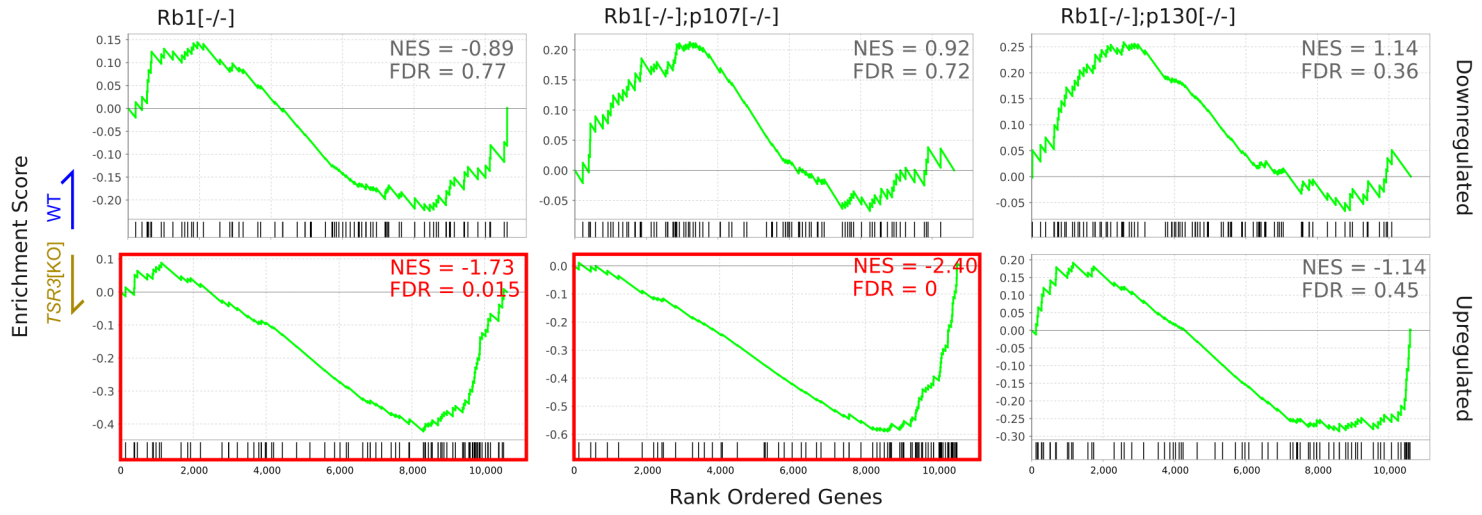
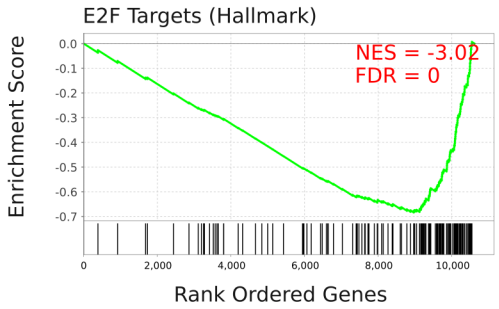

a
HCT116 RNA-seq
WT vs. TSR3[KO/Het]
7
10 545
18
RPKM >0.1
Padj <0.05
b
TSR3[KO 1]
TSR3[KO 3]
WT[3]
WT[1]
WT[2]
TSR3[Het 2]
[2] [2] [1] [3] [3] [1]
c
KEGG Ribosome
RB1[-/-]; p107[-/-] up
log2 Fold-Change
mean RPKM
d i
HCT116 Ribo-seq
Mapped Read Frequency
Read Length (nt)
ii
21-23 nt Read Frequency
WT[1-3]
[Het 2]
[KO 1,3]
e
f
WT vs. TSR3[KO/Het]
668
9 711
760
RPKM >0.1
Padj <0.05
g
KEGG Ribosome
RB1[-/-]; p107[-/-] up
log2 Fold-Change
mean RPKM
h
Met.AUG Occupancy (%)
% in all CDS
i
In-Frame
+1 Frame
+2 Frame
n.s.
P-site Occupancy (%) in CDS
j
KEGG Ribosome
RB1[-/-]; p107[-/-] up
log2 RiboSeq Fold-Change
log2 RNASeq Fold-Change

Figure S6

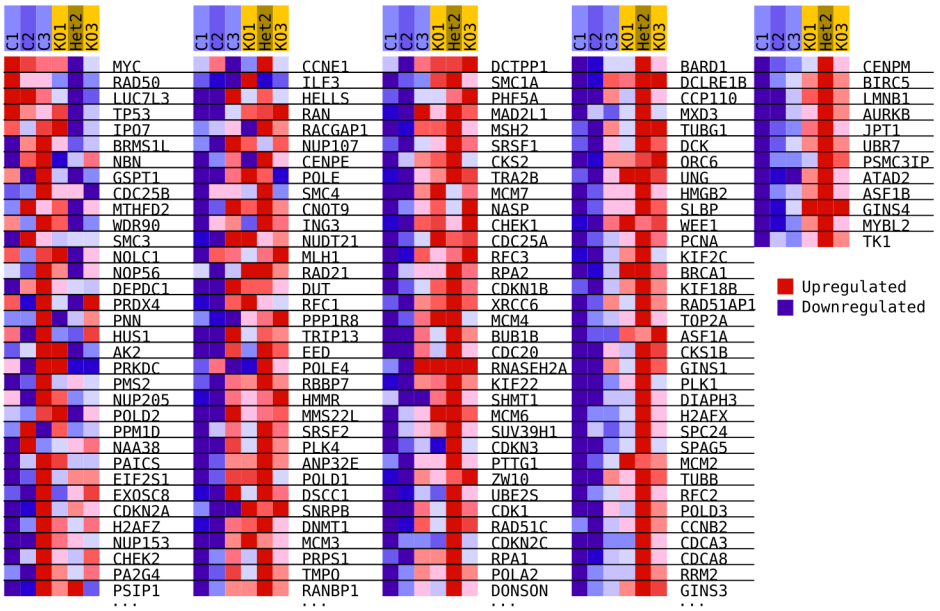
a RNA-seq



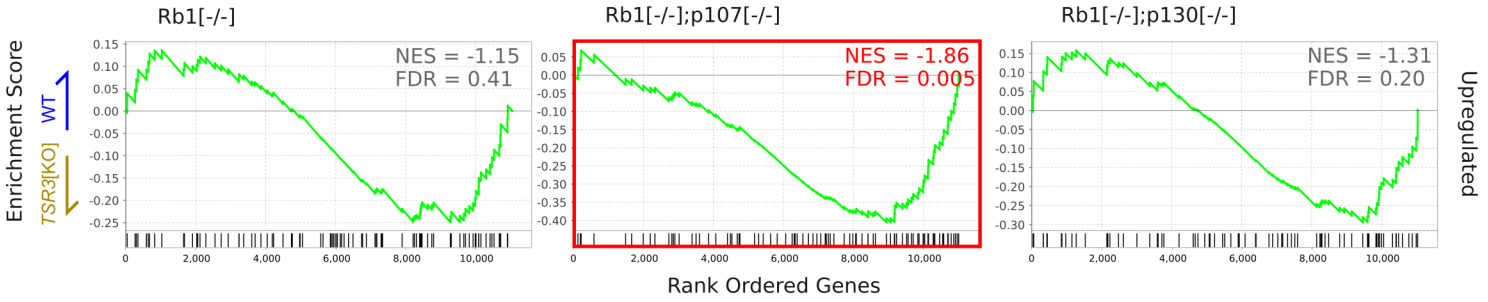
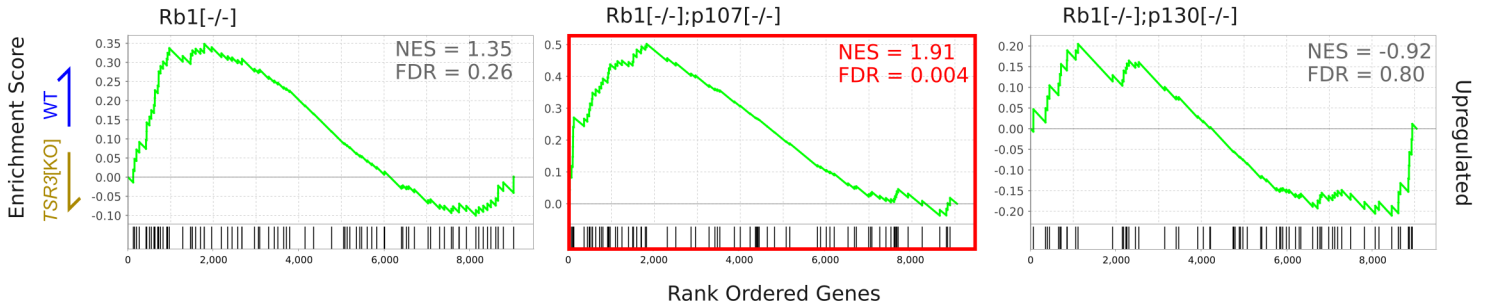
b i.



ii.



c Ribo-seq (Total Translation)



d Translation Efficiency