Alternative splicing through m^{6}A modification at a 3' splice site for SAM synthetase homeostasis

Eichi Watabe, Marina Togo-Ohno, Yuma Ishigami, Shotaro Wani, Keiko Hirota, Mariko Kimura-Asami, Sharmin Hasan, Satomi Takei, Akiyoshi Fukamizu, Yutaka Suzuki, Tsutomu Suzuki and Hidehito Kuroyanagi*

Table of Contents

Appendix Figure S1. Experimental validation of *smg-2*-enriched mRNA isoforms by RT-PCR. ··2

Appendix Figure S2. Alternative splicing of the *sams* genes is dynamically regulated upon feeding and fasting. ·····3

Appendix Figure S3. *sams-5; sams-1* double mutations do not significantly affect ChIP-seq signals for H3K4me2, H3K27me3 or H3K36me3 at the alternatively spliced exons in the *sams-3* and *sams-4* loci. ·····4

Appendix Figure S4. METT-10 is the *C. elegans* orthologue of human METTL16. ·····5

Appendix Figure S5. Specific methylation of the AG dinucleotide in *sams-3*/*sams-4* pre-mRNA by METT-10 *in vitro*. ·····6

Appendix Figure S6. Human METTL16 specifically methylates an adenine base of the AG dinucleotide at the distal 3'SS of *sams-3*/*sams-4* intron 2 *in vitro*. ·····7

Appendix Figure S7. Specific and efficient methylation of *in vitro*-transcribed *sams-3*/*sams-4* and *sams-5* RNAs revealed by LC-MS/MS. ·····8

Appendix Figure S8. Nanopore electric currents at nucleotide positions -4 through +4 relevant to the m^{6}A site. ·····9

Appendix Figure S9. Modeling molecular structure of UAF-2 binding to a 3'SS sequence. ····· 10

Appendix Figure S10. Nucleotide sequence alignment of genomic regions spanning from intron 2 through exon 3 of the *sams* genes in the genus *Caenorhabditis*. ····· 11

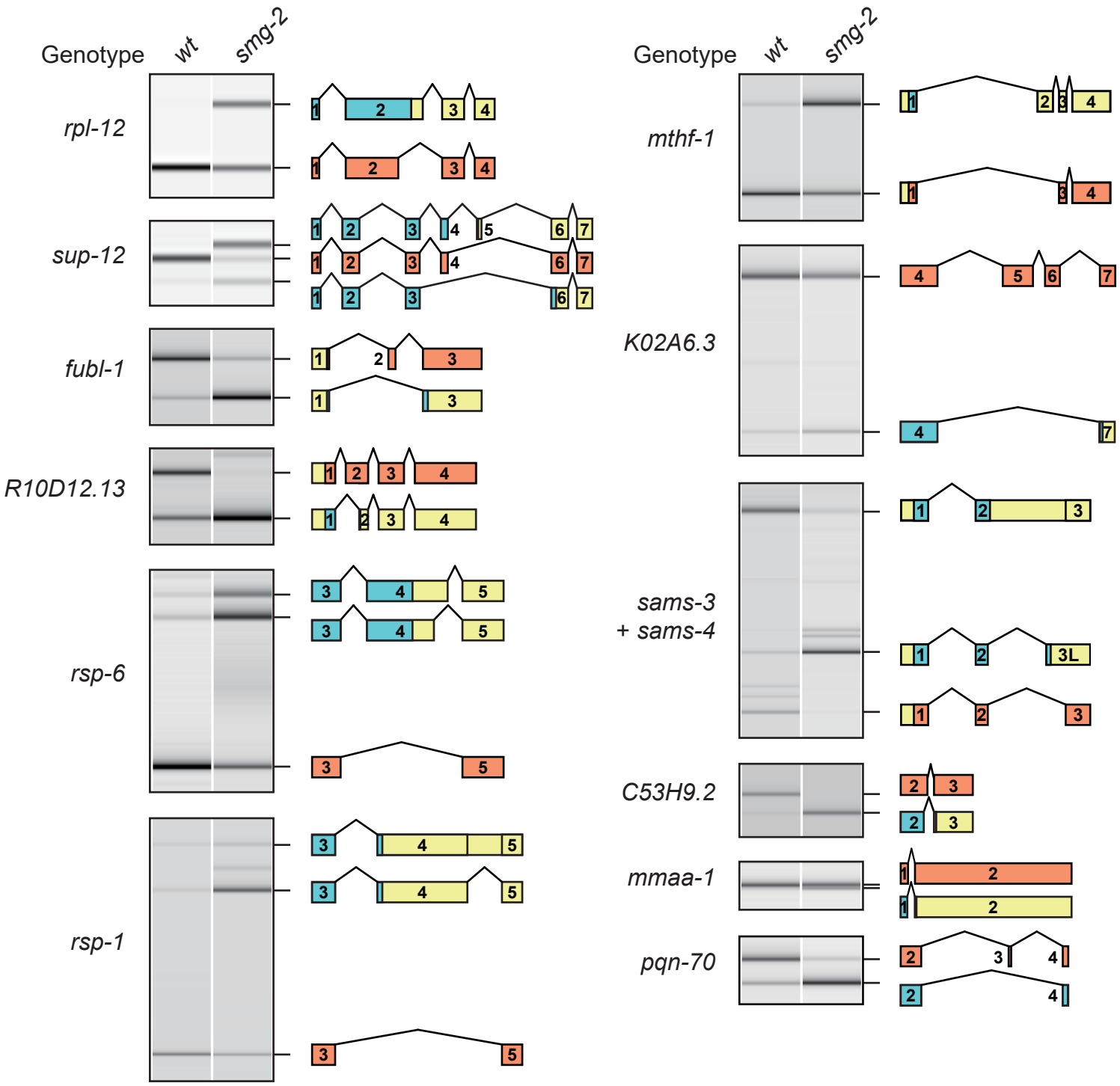
Appendix Figure S11. Base-calling errors in the Nanopore direct RNA sequencing data sets cannot predict the m6A modifications in the *sams* pre-mRNAs. ····· 12

Appendix Figure S12. Anti-SAMS-1, -SAMS-3 and -SAMS-4 antisera specifically recognize their target proteins. ····· 13

Appendix Table S1. Worm strains used in this study. ····· 14

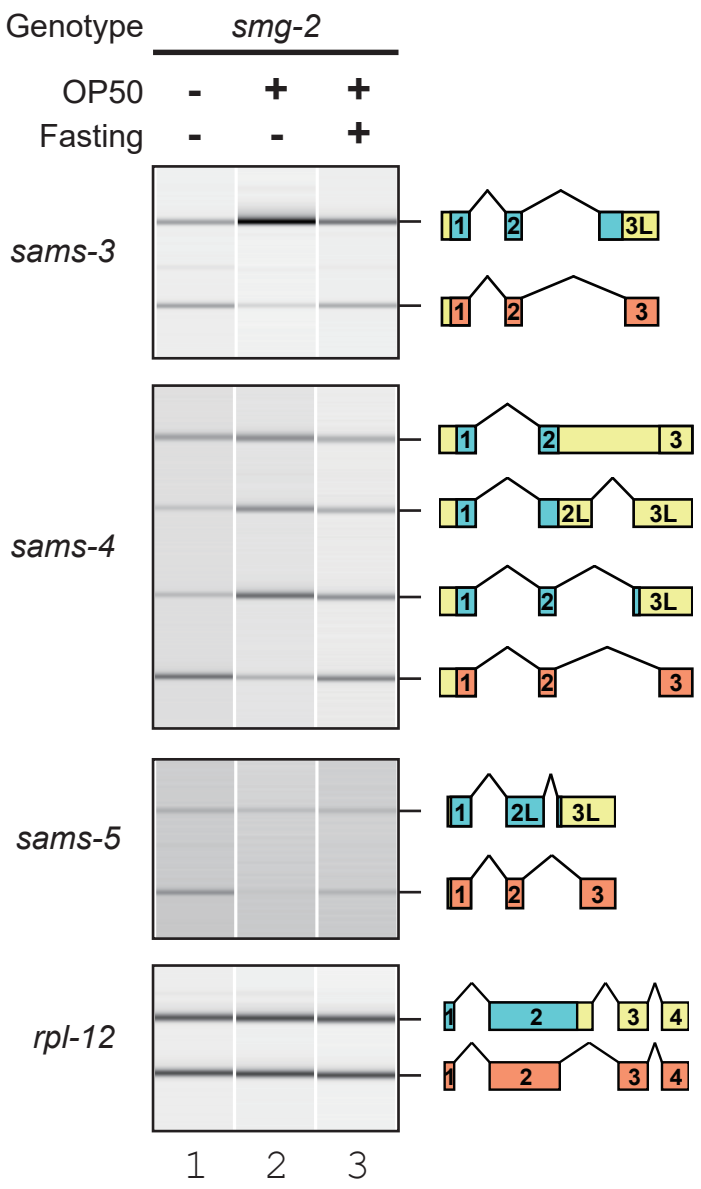
Appendix Table S2. Sequences of primers used to amplify endogenous mRNAs in semi-quantitative RT-PCRs. ····· 15

Appendix Table S3. Sequences of primers used in quantitative PCRs. ····· 16



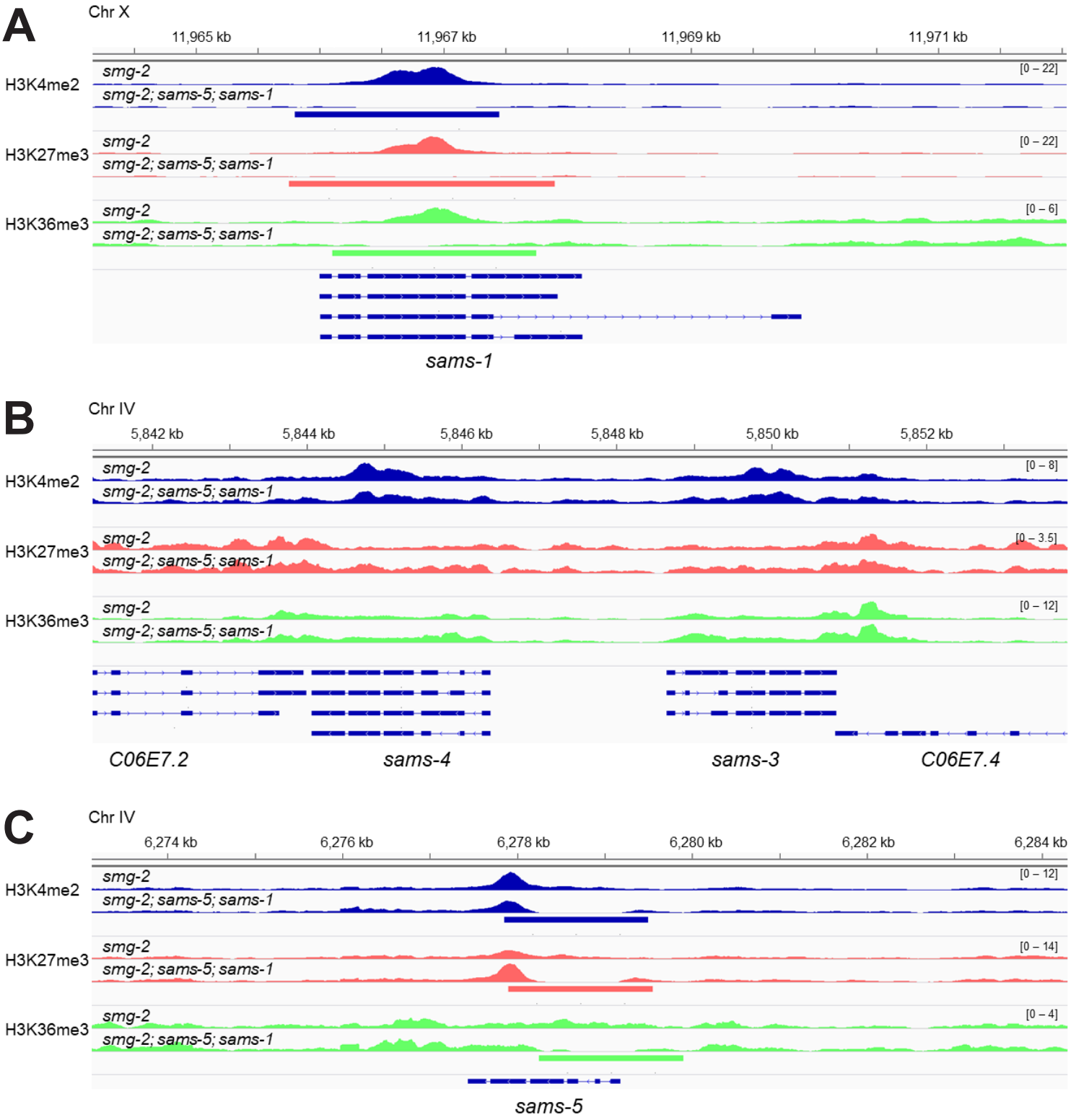
Appendix Figure S1. Experimental validation of *smg-2*-enriched mRNA isoforms by RT-PCR.

Total RNAs were extracted from synchronized L1 larvae of a wild-type strain N2 (*wt*) and an NMD-deficient strain KH1668: *smg-2 (yb979)* and subjected to semi-quantitative RT-PCR, whose products were analyzed by capillary electrophoresis. Schematic structure of each PCR product is indicated on the right. Boxes indicate exons. Open reading frames (ORFs) for full-length and truncated proteins are in orange and cyan, respectively. *rpl-12* was used as a positive control with an NMD isoform. Sequences of the PCR products were confirmed by direct sequencing or cloning and sequencing.



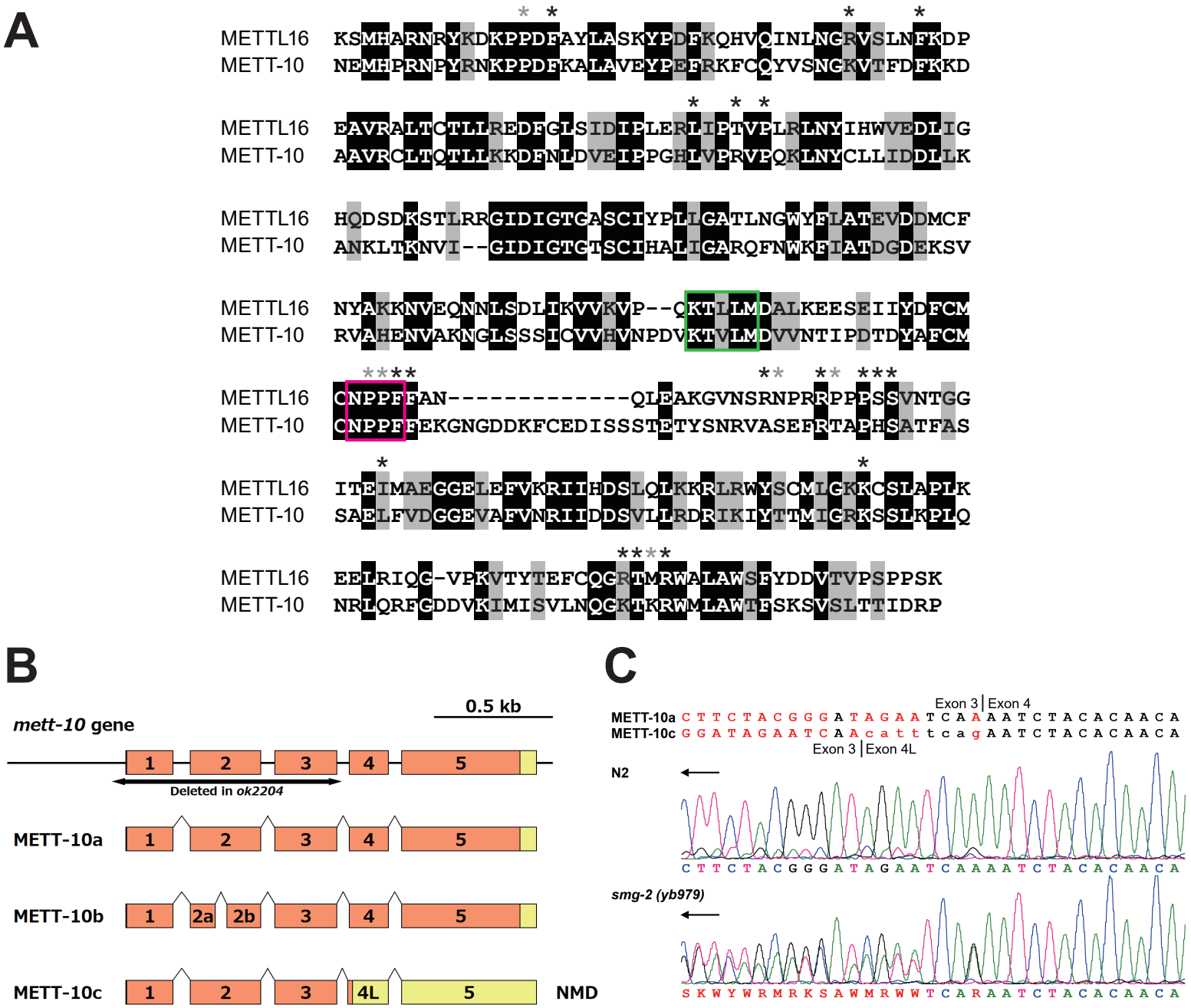
Appendix Figure S2. Alternative splicing of the *sams* genes is dynamically regulated upon feeding and fasting.

Synchronized L1 larvae of an NMD-deficient strain KH1668: *smg-2 (yb979)* were incubated without (lane 1) or with (lane 2) a standard *E. coli* strain OP50 for 3 hours. The fed larvae were further incubated without OP50 for another 3 hours (lane 3). mRNA isoforms were analyzed and presented as in Appendix Figure S1. *rpl-12* was used as an unaffected control with an NMD isoform. Sequences of the PCR products were confirmed by cloning and sequencing.



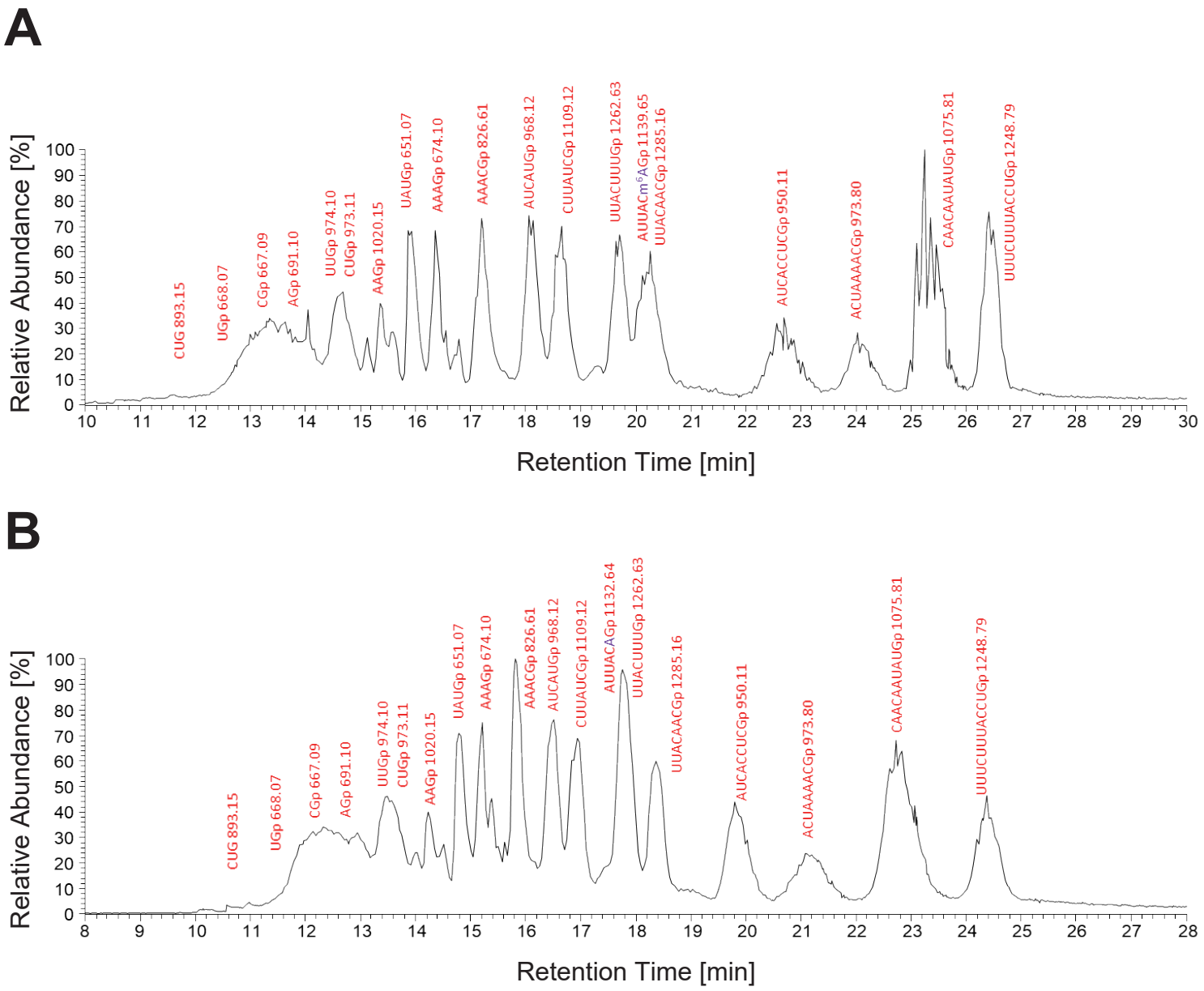
Appendix Figure S3. *sams-5; sams-1* double mutations do not significantly affect ChIP-seq signals for H3K4me2, H3K27me3 or H3K36me3 at the alternatively spliced exons in the *sams-3* and *sams-4* loci.

(A-C) ChIP-seq signals of H3K4me2 (blue), H3K27me3 (red) and H3K36me3 (green) signals at the *sams-1* (A), *sams-3* and *sams-4* (B) and *sams-5* (C) loci in the *smg-2 (yb979)* (top) and *smg-2 (yb979); sams-5 (gk147); sams-1 (ok2946)* (bottom) mutants fed with OP50 in S-complete for 3 hours at 20°C. Note that the ChIP-seq signals are significantly affected in the *sams-1* and *sams-5* loci (indicated with lines below the signals) due to deletions in these genes.



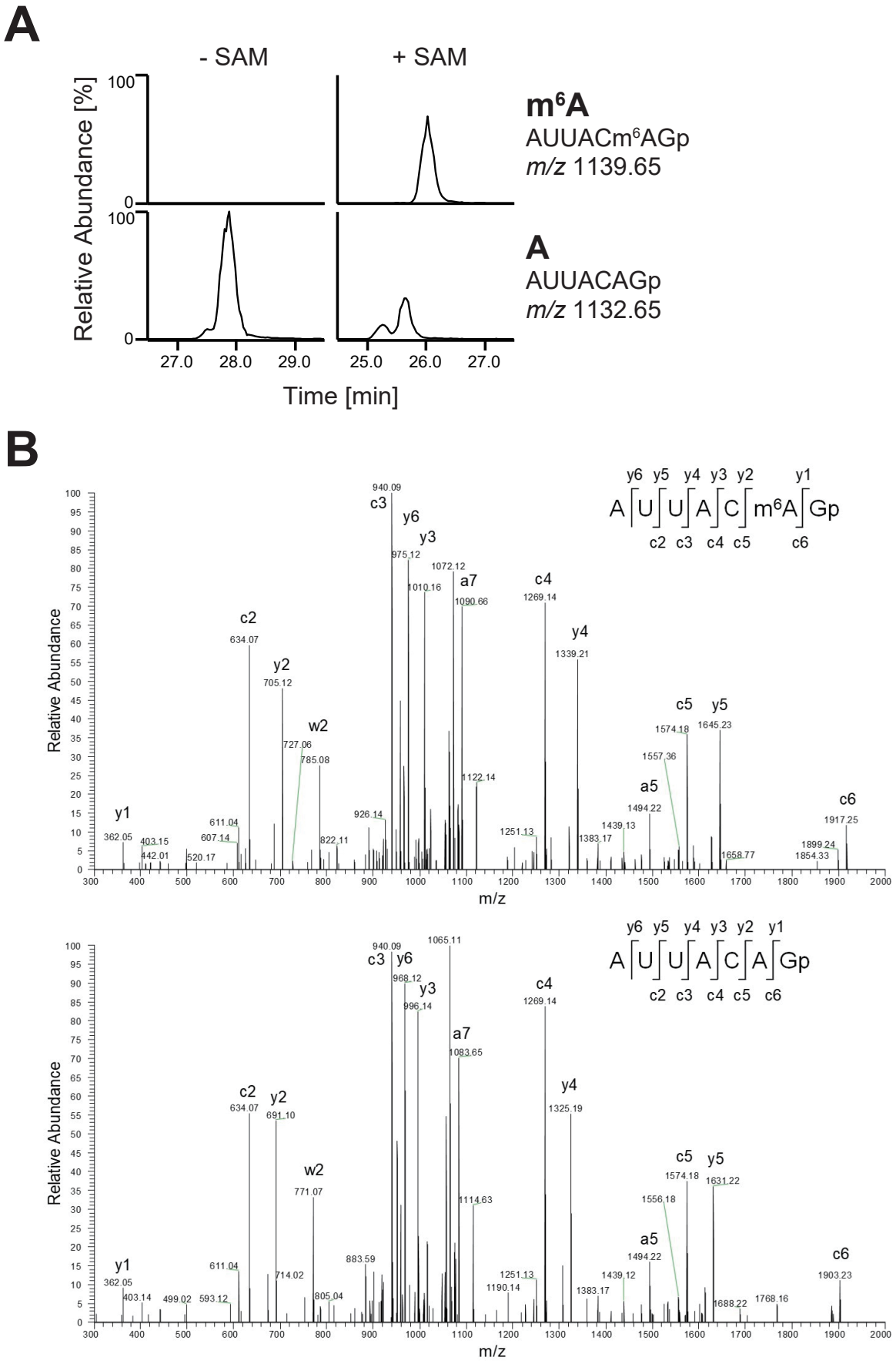
Appendix Figure S4. METT-10 is the *C. elegans* orthologue of human METTL16.

(A) Amino acid sequence alignment of catalytic domains from human METTL16 and *C. elegans* METT-10. Methyltransferase domains from human METTL16 (GenBank Acc. No. NP_076991, aa 5-301) and *C. elegans* METT-10a (GenBank NP_499247, aa 5-315) are aligned by Clustal W. Identical residues are shaded in black. Residues with similar chemical properties are shaded in grey. Black and grey asterisks indicate METTL16 residues whose side chains and main chains, respectively, are directly involved in recognition of the target RNA (Doxtader et al., 2018). Magenta and green boxes indicate the catalytic center and the K-loop, respectively (Doxtader et al., 2018). (B) Schematic structure of the *mett-10* gene and its mRNAs deposited in WormBase or discovered in this study. Note that METT-10c utilizes a cryptic 3'SS for intron 3 that is located 8 nt upstream from the canonical 3'SS. The nucleotide sequence of METT-10c is deposited in the DDBJ/GenBank database under the accession number LC603057. (C) Direct sequencing data of *mett-10* cDNAs amplified from N2 and the *smg-2 (yb979)* mutant.



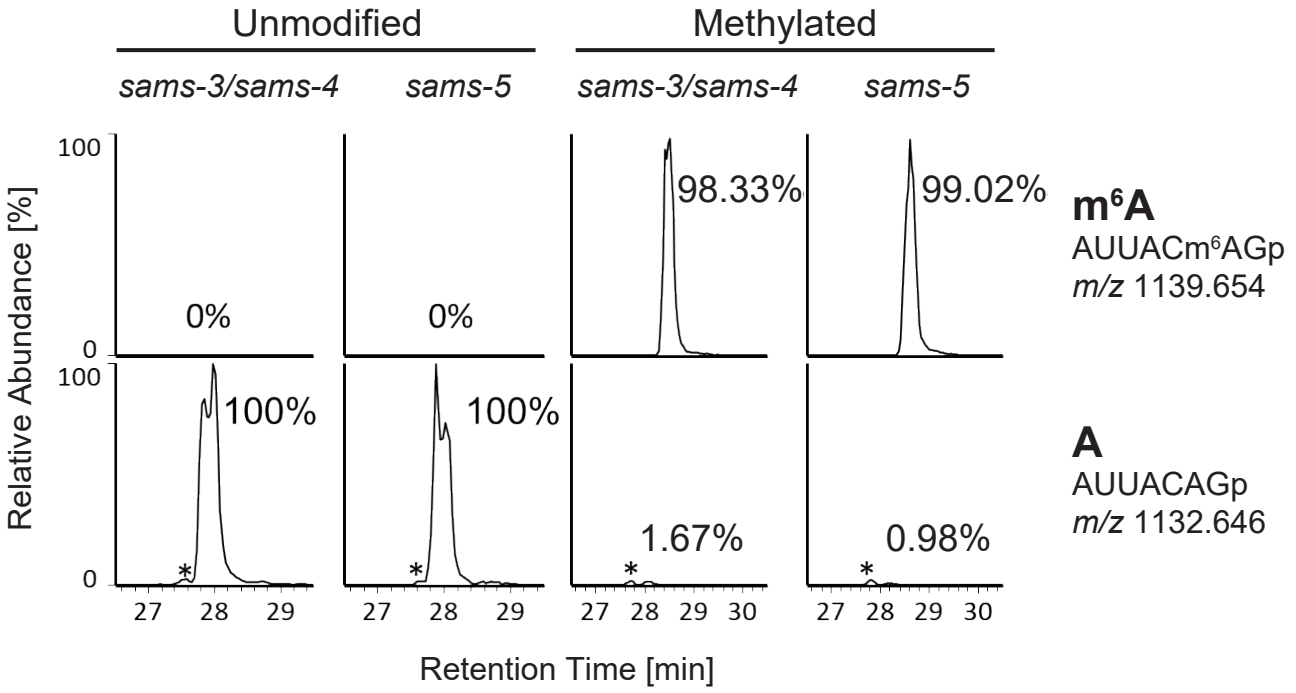
Appendix Figure S5. Specific methylation of the AG dinucleotide in *sams-3*/*sams-4* pre-mRNA by METT-10 *in vitro*.

Total ion chromatograms of RNA fragments from *in vitro*-transcribed *sams-3*/*sams-4* pre-mRNA after *in vitro* methylation with recombinant full-length METT-10 in the presence (A) or absence (B) of 1 mM SAM.



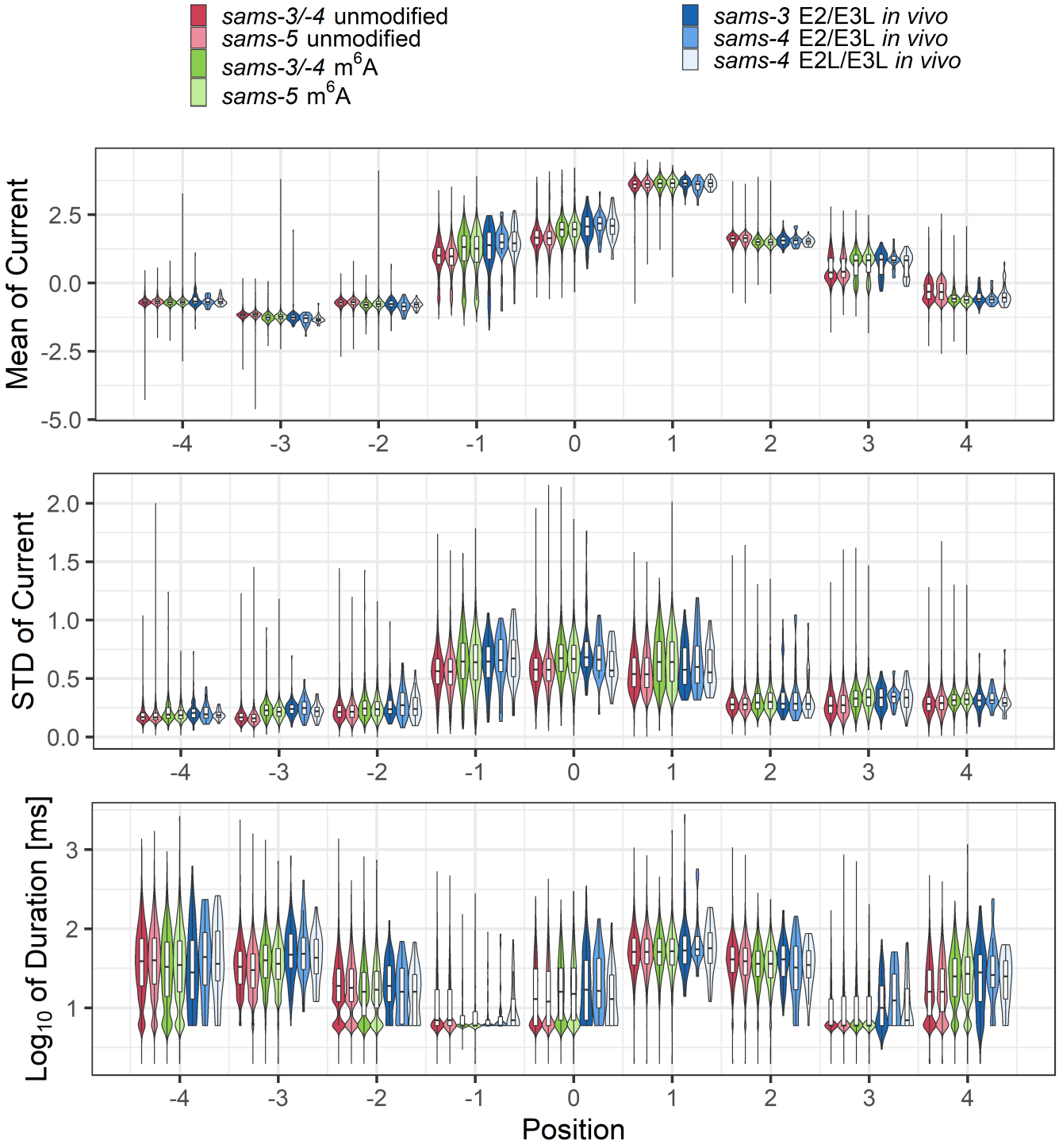
Appendix Figure S6. Human METTL16 specifically methylates an adenine base of the AG dinucleotide at the distal 3'SS of *sams-3*/*sams-4* intron 2 *in vitro*.

(A) Mass chromatograms of RNA fragments to detect m^{6}A and unmodified A after *in vitro* incubation of *sams-3*/*sams-4* pre-mRNA with recombinant METTL16 methyltransferase domain as in Figure 5A. (B) CID spectra of the methylated (top) and unmethylated (bottom) RNA fragments to map the position of m^{6}A.



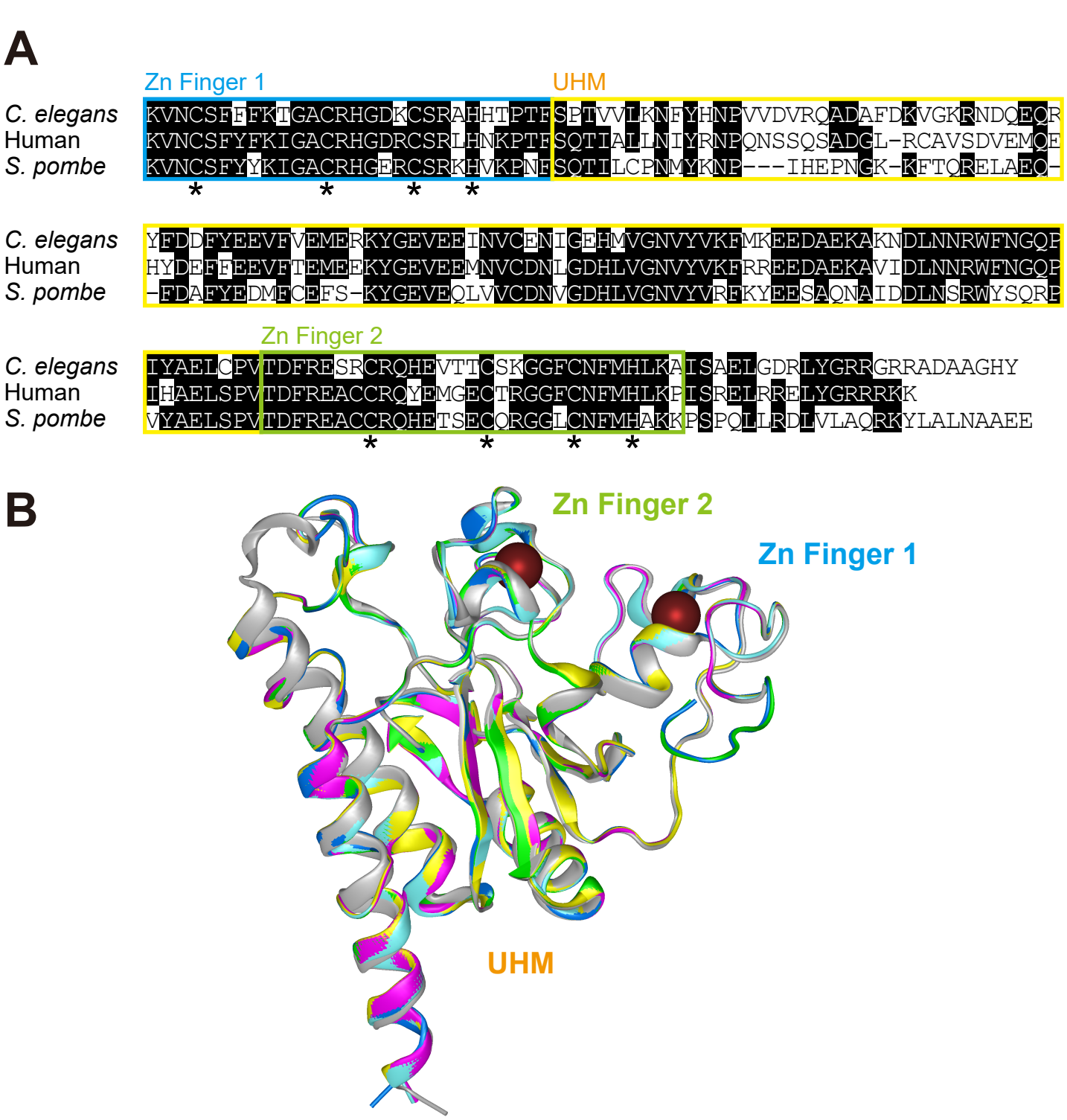
Appendix Figure S7. Specific and efficient methylation of *in vitro*-transcribed *sams-3*/*sams-4* and *sams-5* RNAs revealed by LC-MS/MS.

Mass chromatograms of RNA fragments to quantify m^6A and unmodified A in AUUAC<u>A</u>Gp fragments after incubation of *in vitro*-transcribed, barcoded *sams-3*/*sams-4* and *sams-5* RNAs with recombinant human METTL16 methyltransferase domain as in Appendix Figure S6A. These four RNAs were pooled and analyzed by Nanopore direct RNA sequencing in Figure 6C.



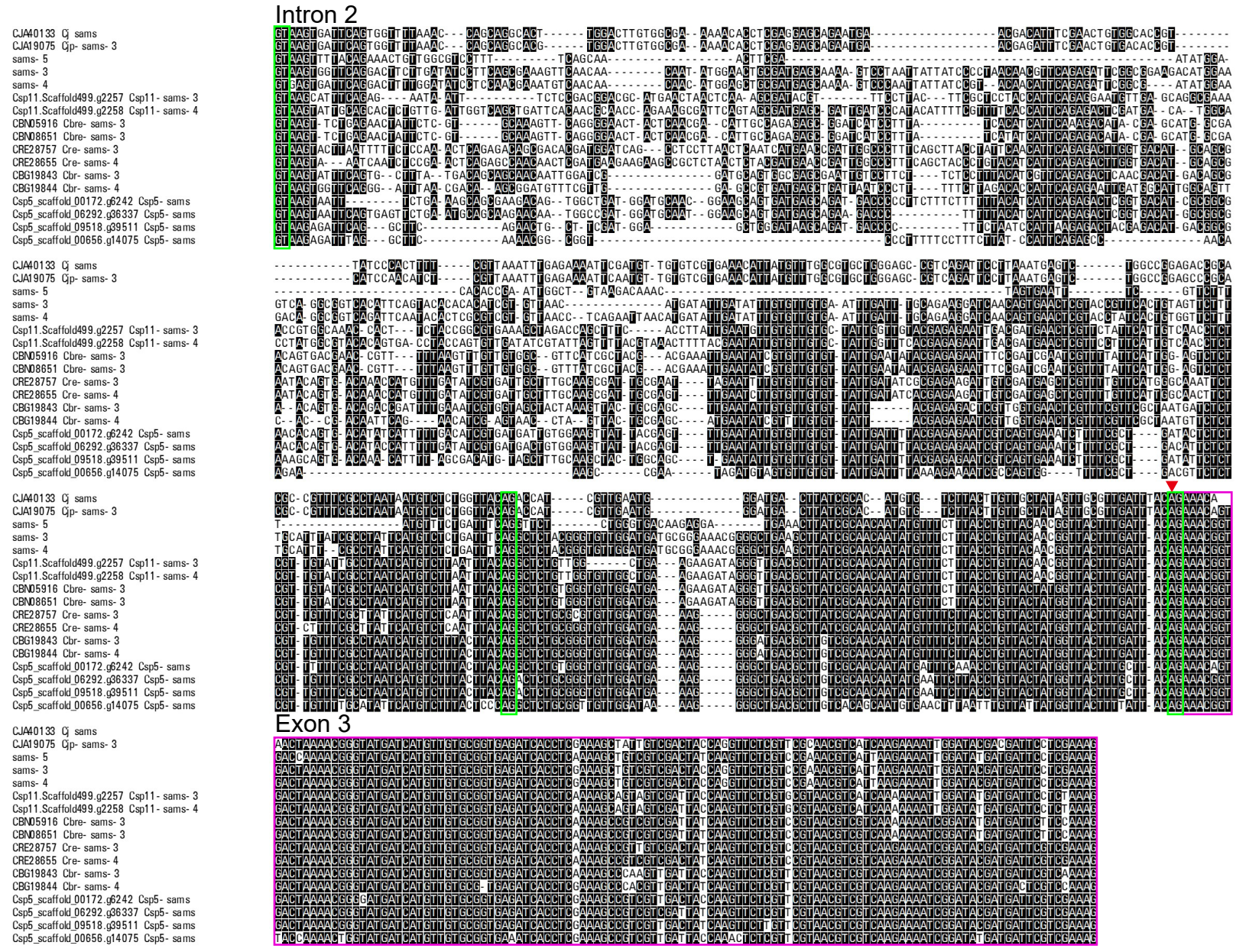
Appendix Figure S8. Nanopore electric currents at nucleotide positions -4 through +4 relevant to the m⁶A site.

Distribution of mean (top), standard deviation (middle) and duration time (bottom) of normalized Nanopore currents for *in vitro*-transcribed unmodified *sams-3*/*sams-4* and *sams-5* RNAs (unmodified), those methylated with the METTL16 methyltransferase domain *in vitro* (m⁶A) and endogenous *sams-3* and *sams-4* mRNA isoforms (*in vivo*). Color codes are indicated at the top.



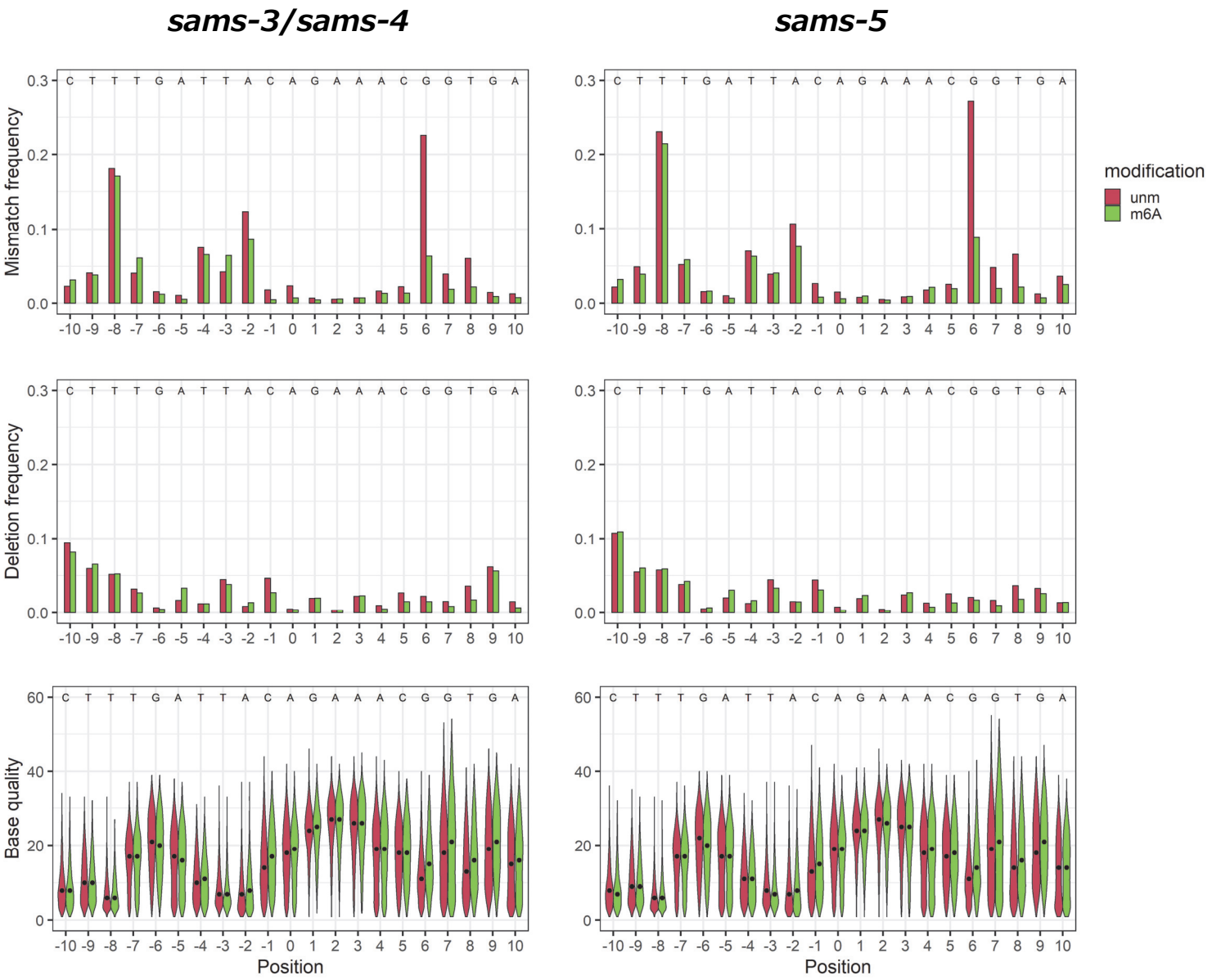
Appendix Figure S9. Modeling molecular structure of UAF-2 binding to a 3'SS sequence.

(A) Amino acid sequence alignment of CCCH-type Zn finger domains and a U2AF homology motif (UHM) domain among U2AF35 orthologues, *C. elegans* UAF-2 (GenBank Acc No. NP_503036, aa 23-208), human U2AF35 (GenBank NP_001307575, aa 15-192) and *S. pombe* U2AF23 (GenBank NP_594945, aa 15-194). Identical residues are shaded. The Zn finger 1, UHM and Zn finger 2 domains are boxed in blue, yellow and green, respectively. Asterisks indicate cysteine and histidine residues coordinating to zinc ions. (B) Models of UAF-2 three-dimensional structure based on amino acid sequence homology to U2AF23 from *S. pombe*. Five modeled structures are in different colors and the crystal structure of *S. pombe* U2AF23 (Yoshida et al., 2020) in complex with U2AF59 (not shown) and 5'-UAGGU-3' (not shown) is in grey. Red spheres indicate zinc ions.

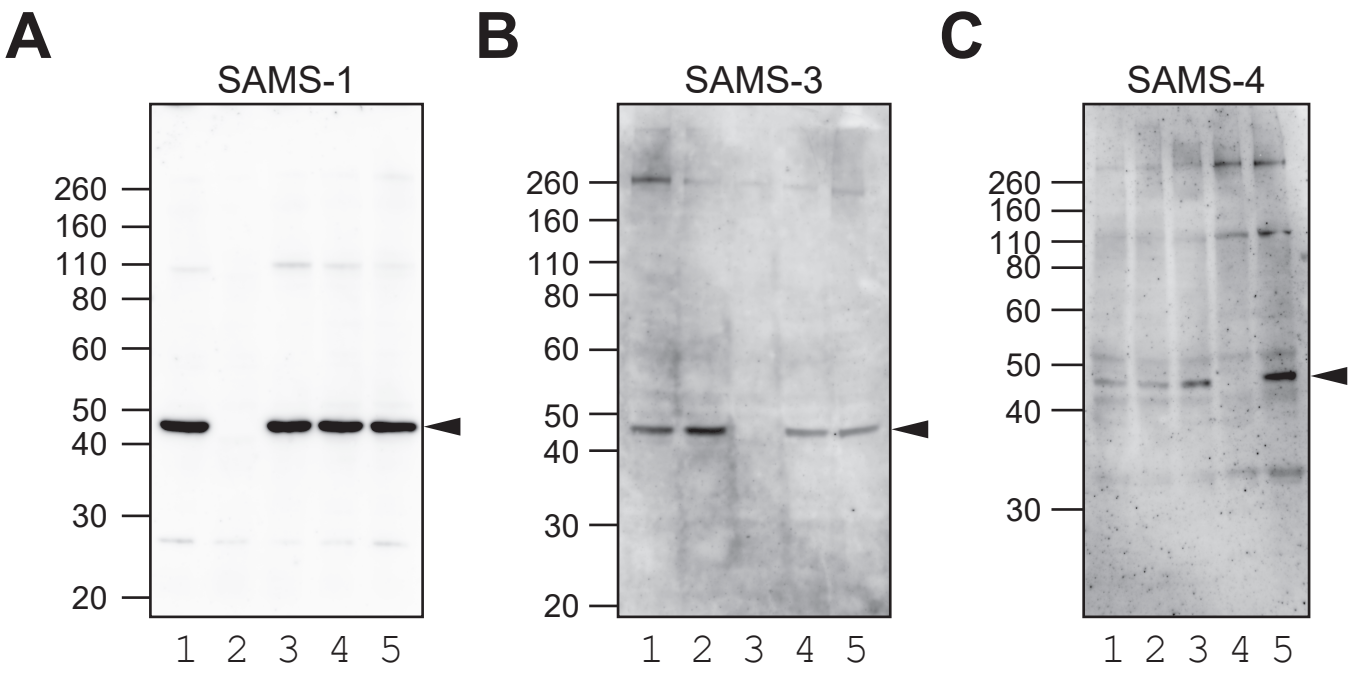


Appendix Figure S10. Nucleotide sequence alignment of genomic regions spanning from intron 2 through exon 3 of the *sams* genes in the genus *Caenorhabditis*.

The sequences were aligned with Clustal V algorithm. Conserved nucleotides are highlighted in black. Exons are boxed in magenta. Canonical 5'SSs as well as proximal/unproductive and distal/productive 3'SSs are boxed in green. A red arrowhead indicates the m^{6}A modification sites in *C. elegans*.



Appendix Figure S11. Base-calling errors in the Nanopore direct RNA sequencing data sets cannot predict the m^{6}A modifications in the *sams* pre-mRNAs.
We tried to predict the m^{6}A modifications in our unmodified and *in vitro*-methylated *sams-3*/*sams-4* (left) and *sams-5* (right) pre-mRNAs with the "*EpiNano*" algorithm (Liu *et al*., 2019) and the mismatch frequencies (top), deletion frequencies (middle) and base qualities (bottom) at nucleotide positions -10 through +10 relevant to the m^{6}A site are indicated. Medians of the base quality scores are indicated with dots. Number of reads used in the panels are 7,700 and 2,796 for the unmodified and m^{6}A-modified *sams-3*/*sams-4* RNAs, respectively, and 12,411 and 3,958 for the unmodified and m^{6}A-modified *sams-5* RNAs, respectively. Note that the error rates for the m^{6}A-modified RNAs were rather less than those for the unmodified RNAs and therefore cannot be used to predict the m^{6}A modification sites.



Appendix Figure S12. Anti-SAMS-1, -SAMS-3 and -SAMS-4 antisera specifically recognize their target proteins.

Western blot analysis of SAMS-1 (A), SAMS-3 (B) and SAMS-4 (C) proteins in young adult worm lysates from *smg-2 (yb979)* (lane 1), *smg-2 (yb979); sams-1 (ok2946)* (lane 2), *smg-2 (yb979); sams-3 (tm4237)* (lane 3), *smg-2 (yb979); sams-4 (tm4235)* (lane 4) and *smg-2 (yb979); sams-5 (gk147)* (lane 5) mutants.

Appendix Table S1. Worm strains used in this study.

Strain	Genotype
KH1668	*smg-2 (yb979) I* (Outcrossed 6 times)
KH2411	*smg-2 (yb979) I; sams-5 (gk147) IV*
KH2416	*smg-2 (yb979) I; sams-4 (tm4235) IV*
KH2417	*smg-2 (yb979) I; sams-3 (tm4237) IV*
KH2418	*smg-2 (yb979) I; sams-1 (ok2946) X*
KH2421	*smg-2 (yb979) I; sams-5 (gk147) IV; sams-1 (ok2946) X*
KH2546	*smg-2 (yb979) I; mett-10 (ok2204) III*
N2	*Wild type*
VC1743	*mett-10 (ok2204) III*
VC2428	*sams-1 (ok2946) X*

Appendix Table S2. Sequences of primers used to amplify endogenous mRNAs in semi-quantitative RT-PCRs.

Gene	Direction	Sequence
C53H9.2	Forward	5′-GAACTCCGCTATTGTCGTCC-3′
C53H9.2	Reverse	5′-AACTCTCCAGAGCTCTCTCC-3′
fubl-1	Forward	5′-CACCATGGAAGGATTTAACCCGCAG-3′
fubl-1	Reverse	5′-TGAACACGACAACCAGCCTTG-3′
K02A6.3	Forward	5′-TCAGTTCTCAAGACCGACCA-3′
K02A6.3	Reverse	5′-CTGCTCTGCAAGCTCCATTG-3′
mett-10	Forward	5′-CACCATGTCACAAAACAACGAGATGCA-3′
mett-10	Reverse	5′-TGTCCATGCCAACATCCAGC-3′
mmaa-1	Forward	5′-GCTTCGGGTATCCAGACTCA-3′
mmaa-1	Reverse	5′-GTCATCCCAATGAACACGAG-3′
mthf-1	Forward	5′-GATTCTCTGTCCACAATGCCA-3′
mthf-1	Reverse	5′-GAAAGACGTTCAACACGCTC-3′
pqn-70	Forward	5′-TCACCAGCCTCCAATGATGCA-3′
pqn-70	Reverse	5′-CCAATAGCTCCTTGTCCCAC-3′
R10D12.13	Forward	5′-GAAGAGGAACTGACAGAAGATG-3′
R10D12.13	Reverse	5′-ACACAGTTGGACGAAGATGG-3′
rpl-12	Forward	5′-CACCATGCCACCAAAGTTCGACCCA-3′
rpl-12	Reverse	5′-TCATTGAGCTGGGATCTCGAT-3′
rsp-1	Forward	5′-TCGCTACGGTCGTCCATACA-3′
rsp-1	Reverse	5′-TGGTCTCTTATGAGCTTCGG-3′
rsp-6	Forward	5′-GAATCTGTGGTGTCCGTGCT-3′
rsp-6	Reverse	5′-CTGCGTTCACGACTGCGATC-3′
sams-3 + *sams-4*	Forward	5′-CACCTCCGAATCTGTCTCTG-3′
sams-3 + *sams-4*	Reverse	5′-TGATCTCACCGCACAACATG-3′
sams-3	Forward	5′-GTTTCAACCTTACAACCAAA-3′
sams-4	Forward	5′-GACGCTTTCAAACATCCATCA-3′
sams-5	Forward	5′-ACAAAGCTTCCCCAGCAAGT-3′
sams-3, *sams-4*, *sams-5*	Reverse	5′-CGGACGAGAACYTGRTAGTC-3′
sup-12	Forward	5′-CACCATGTACGGCCAAGTCCAAGAT-3′
sup-12	Reverse	5′-CTAGACACGTTGATGCTCAAGCTG-3′

Underlines indicate CACC sequences for directional TOPO cloning (Invitrogen).

Appendix Table S3. Sequences of primers used in quantitative PCRs.

Gene	Isoform	Direction	Sequence
acs-2	-	Forward	5′-CGCAAAGCTGGAAGACATGG-3′
acs-2	-	Reverse	5′-GTTGTTCAGCCCGAAATGGG-3′
eef-1A.1	-	Forward	5′-CACTCCGTAGCAGCCATGGG-3′
eef-1A.1	-	Reverse	5′-CTCCTGAGCCTCCTTCTCG-3′
fat-7	-	Forward	5′-TTGCCATCACAAGTGGACTGA-3′
fat-7	-	Reverse	5′-GATAGGTCGAGTTTGCCTCC-3′
sams-1	-	Forward	5′-CAGAAATGTCCTCCAAATTCC-3′
sams-1 *sams-4* *sams-5*	- E2L/E3L Unproductive	Reverse	5′-CGGACGAGAACYTGRTAGTC-3′
sams-2	-	Forward	5′-GTTTCAACCTTCCAACTG-3′
sams-2 *sams-3* *sams-4* *sams-5*	- Productive Productive Productive	Reverse	5′-AGTCACCGTTTCACAWGC-3′
sams-3	Total	Forward	5′-TCTGTCGCCGTTGTCTTGTT-3′
sams-3	Total	Reverse	5′-CTTCAAATCGAGATCCTTAAT-3′
sams-3	Productive Unproductive	Forward	5′-GTTTCAACCTTACAACCAAA-3′
sams-3 *sams-4*	Unproductive E2/E3L	Reverse	5′-ACCCGTAGAGCCACAWGC-3′
sams-4	Total	Forward	5′-CAAGTTTCATACGCCATC-3′
sams-4	Total	Reverse	5′-CTTCAAGTCAWAGATTCTTGAT-3′
sams-4	Productive E2/E3L	Forward	5′-GACGCTTTCAAACATCCATCA-3′
sams-4	E2L/E3L	Forward	5′-GAAGACAGGCGGCTCTACG-3′
sams-5	Total Productive	Forward	5′-ACAAAGCTTCCCCAGCAAGT-3′
sams-5	Total	Reverse	5′-GCATCGAGGACCGCATCG-3′
sams-5	Unproductive	Forward	5′-CACCGAATTGGCTGTTCTC-3′