

**Editorial summary**

Nanopore sequencing detects pseudouridine and 2'-O-methylation modifications in cellular RNAs

**Quantitative profiling of pseudouridylation dynamics in native RNAs with nanopore sequencing**

Oguzhan Begik<sup>1,2,3,#</sup>, Morghan C Lucas<sup>1,4,#</sup>, Leszek P Pryszcz<sup>1,5</sup>, Jose Miguel Ramirez<sup>1</sup>, Rebeca Medina<sup>1</sup>, Ivan Milenkovic<sup>1,4</sup>, Sonia Cruciani<sup>1,4</sup>, Huanle Liu<sup>1</sup>, Helaine Grazielle Santos Vieira<sup>1</sup>, Aldema Sas-Chen<sup>6</sup>, John S Mattick<sup>3</sup>, Schraga Schwartz<sup>6</sup> and Eva Maria Novoa<sup>1,4,7\*</sup>

<sup>1</sup>*Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, Dr. Aiguader 88, Barcelona 08003, Spain*

<sup>2</sup>*Garvan Institute of Medical Research, Darlinghurst, NSW, 2010, Australia*

<sup>3</sup>*UNSW Sydney, Kensington NSW, 2052, Australia*

<sup>4</sup>*Universitat Pompeu Fabra (UPF), Barcelona, Spain*

<sup>5</sup>*International Institute of Molecular and Cell Biology, 4 Ks. Trojdena Street, 02-109 Warsaw, Poland*

<sup>6</sup>*Weizmann Institute of Science, Rehovot, IL*

<sup>7</sup>*Lead Contact*

<sup>#</sup>*These authors contributed equally*

\* Correspondence to: Eva Maria Novoa ([eva.novoa@crg.eu](mailto:eva.novoa@crg.eu))

35 Nanopore RNA sequencing shows promise as a method for discriminating and identifying  
36 different RNA modifications in native RNA. Expanding on the ability of nanopore sequencing  
37 to detect N6-methyladenosine (m6A), we show that other modifications, in particular  
38 pseudouridine ( $\Psi$ ) and 2'-O-methylation (Nm), also result in characteristic base-calling 'error'  
39 signatures in the nanopore data. Focusing on  $\Psi$  modification sites, we detect known and  
40 uncover previously unreported  $\Psi$  sites in mRNAs, ncRNAs and rRNAs, including a Pus4-  
41 dependent  $\Psi$  modification in yeast mitochondrial rRNA. To explore the dynamics of  
42 pseudouridylation, we treat yeast cells with oxidative, cold and heat stresses and detect heat-  
43 sensitive  $\Psi$ -modified sites in snRNAs, snoRNAs and mRNAs. Finally, we develop a software,  
44 *nanoRMS*, that estimates per-site modification stoichiometries by identifying single-molecule  
45 reads with altered current intensity and trace profiles. This work demonstrates that Nm and  $\Psi$   
46 RNA modifications can be detected in cellular RNAs and that  $\Psi$  RNA can be identified in a  
47 quantitative manner by nanopore sequencing of native RNA.

48  
49  
50  
51

52

53 RNA modifications are chemical moieties that decorate RNA molecules, expanding their lexicon. By  
54 coupling antibody immunoprecipitation or chemical probing with next-generation sequencing (NGS),  
55 transcriptome-wide maps of several RNA modifications have been constructed, including N6-  
56 methyladenosine (m<sup>6</sup>A)<sup>1,2</sup>, pseudouridine (Ψ)<sup>3-6</sup>, 5-methylcytosine (m<sup>5</sup>C)<sup>7,8</sup>, 5-  
57 hydroxymethylcytosine (hm<sup>5</sup>C)<sup>9</sup>, 1-methyladenosine (m<sup>1</sup>A)<sup>10,11</sup>, N3-methylcytosine (m<sup>3</sup>C)<sup>12</sup>, N4-  
58 acetylcytosine (ac<sup>4</sup>C)<sup>13,14</sup> and 7-methylguanosine (m<sup>7</sup>G)<sup>15,16</sup>. These studies have revealed that RNA  
59 modifications play a pivotal role in many cellular processes, including regulation of cellular fate<sup>17</sup>, sex  
60 determination<sup>18</sup> and cellular differentiation<sup>19</sup>.

61 Despite these advances, a fundamental challenge in the field is the lack of a generic approach for  
62 mapping diverse RNA modification types simultaneously<sup>20-23</sup>. Currently, customized protocols must  
63 be optimized for each RNA modification type, leading to experimental designs in which the RNA  
64 modification type to be studied is chosen beforehand, hindering the ability to characterize the plasticity  
65 of the epitranscriptome in a systematic and unbiased manner. Moreover, even in those cases where a  
66 selective antibody or chemical is available, NGS-based methods are often not quantitative, have high  
67 false positive rates<sup>21</sup>, are inconsistent when using distinct antibodies<sup>24</sup>, are unable to produce maps  
68 for highly repetitive regions, cannot provide information regarding the co-occurrence of distant  
69 modifications in same transcripts, do not provide isoform-specific information, and require multiple  
70 ligation steps and extensive PCR amplification during the library preparation, introducing undesired  
71 biases in the sequencing data<sup>25</sup>.

72 A promising alternative to NGS-based technologies is the direct RNA sequencing platform developed  
73 by Oxford Nanopore Technologies (ONT), which has the potential to detect virtually any given RNA  
74 modification present in native RNA molecules<sup>20,26,27</sup>. Several studies have shown that RNA  
75 modifications can be identified using this technology<sup>28-30</sup>; however, most efforts have been so far  
76 limited to the detection of m<sup>6</sup>A modifications<sup>28-34</sup>, and it is largely unknown whether other RNA  
77 modifications may be distinguishable from their unmodified counterparts. Thus, a systematic approach  
78 that can map and quantify diverse RNA modifications simultaneously is currently missing.

79

80 Here, we examine the *S. cerevisiae* coding and non-coding transcriptome at single molecule  
81 resolution using native RNA nanopore sequencing. We find that most RNA modifications cause  
82 systematic base-calling errors, and that the signature of these base-calling 'errors' can be used to  
83 identify the underlying RNA modification type. For example, we find that pseudouridine typically  
84 appears in the form of U-to-C mismatches, whereas m<sup>5</sup>C modifications appear in the form of  
85 insertions. We then exploit the identified signatures to *de novo* predict RNA modifications in rRNAs,  
86 identifying two previously unreported Ψ modifications in mitochondrial rRNA, which we confirm using  
87 CMC-probing coupled to nanopore sequencing (nanoCMC-seq). We demonstrate that one of these Ψ  
88 modifications (15s:Ψ854) is placed by the enzyme Pus4, which was previously thought to  
89 pseudouridylate only mRNAs and tRNAs<sup>4</sup>. Moreover, we show that once RNA modifications have  
90 been accurately predicted using base-calling 'errors', the stoichiometry of a given Ψ- or Nm-modified

91 site can be estimated by clustering per-read features (current intensities and trace) of the modified  
92 regions.

93

94 We then explore the dynamics of RNA modifications present in non-coding RNAs. It has been  
95 proposed that differential rRNA modifications may constitute a source of ribosomal heterogeneity<sup>35–37</sup>.  
96 Indeed, previous studies have shown that temperature changes affect rRNA pseudouridylation levels,  
97 suggesting that cells may be able to generate compositionally distinct ribosomes in response to  
98 environmental cues<sup>4,38,39</sup>. Similarly, alterations in the stoichiometry of 2'-O-methylation (Am, Cm, Gm,  
99 Um)<sup>40–42</sup> and pseudouridylation ( $\Psi$ )<sup>35–37</sup> can affect translation initiation of mRNAs containing internal  
100 ribosome entry sites (IRES)<sup>43,44</sup>. Here we re-examine this question using direct RNA sequencing, and  
101 characterize the RNA modification dynamics in rRNAs, snRNAs and snoRNAs upon a battery of  
102 environmental cues, translational repertoires and genetic strains. Contrary to expectations, we find  
103 that none of the environmental stresses tested lead to significant changes in the ribosomal  
104 epitranscriptome. By contrast, our method does recapitulate previously reported heat-dependent  $\Psi$   
105 snRNA modifications, as well as identifies previously unreported heat-sensitive sites in snRNAs and  
106 snoRNAs.

107

108 Finally, we develop an algorithm, *nanoRMS*, which we demonstrate can predict  $\Psi$  RNA modifications  
109 *de novo*, and estimate the stoichiometry of modification both in highly- and lowly-modified  $\Psi$  and Nm  
110 sites across diverse types of RNA molecules, including rRNAs, sn/snoRNAs and mRNAs. Our  
111 approach recapitulates known Pus1-dependent, Pus4-dependent and heat stress-dependent mRNA  
112 sites, as well as reveals  $\Psi$  mRNA sites that had not been previously reported. Altogether, our work  
113 establishes a framework for the study of RNA modification dynamics using direct RNA nanopore  
114 sequencing, opening avenues to study the plasticity of the epitranscriptome at single molecule  
115 resolution.

116

117

## RESULTS

### RNA modification detection depends on base-calling and mapping algorithms

Previous studies have shown that m<sup>6</sup>A RNA modifications can be detected in the form of non-random base-calling ‘errors’ in direct RNA sequencing datasets<sup>29–33</sup>. However, it is unclear how these ‘errors’ may vary with the choice of base-calling and mapping algorithms, and consequently, affect the ability to detect RNA modifications. Here, we compared the performance of commonly used base-calling and mapping algorithms on *in vitro* transcribed RNA sequences that contained all possible combinations of 5-mers, referred to as ‘curlcakes’ (CCs)<sup>29</sup>, that included: (i) unmodified nucleosides (UNM), (ii) N6-methyladenosine (m<sup>6</sup>A), (iii) pseudouridine (Ψ), (iv) N5-methylcytosine (m<sup>5</sup>C), and (v) N5-hydroxymethylcytosine (hm<sup>5</sup>C) (**Figure 1A**). In addition, a sixth dataset containing unmodified short RNAs (UNM-S), with median length of 200 nucleotides, was included in the analysis to assess the effect of input sequence length in base-calling (see *Methods*). Each dataset was base-called with two distinct algorithms (*Albacore* and *Guppy*), and using two different versions for each of them, namely: (i) *Albacore* version 2.1.7 (AL 2.1.7); (ii) its latest version, *Albacore* 2.3.4 (AL 2.3.4); (iii) *Guppy* 2.3.1 (GU 2.3.1); and (iv) a more recent version of the latter base-caller, *Guppy* 3.0.3 (GU 3.0.3), which employs a flip-flop algorithm. We found that the latest version of *Albacore* (2.3.4) base-called 100% of sequenced reads in all 6 datasets, whereas its previous version did not (average of 90.8%) (**Figure 1B**). By contrast, both versions of *Guppy* (2.3.1 and 3.0.3) produced similar results in terms of percentage of base-called reads (99.96% and 100%, respectively) (**Table S1**).

We then assessed whether the choice of mapper might affect the ability to detect RNA modifications. To this end, we employed two commonly used long-read mappers, *minimap2*<sup>45</sup> and *GraphMap*<sup>46</sup>, using either ‘default’ or ‘sensitive’ parameter settings (see *Methods*). Notably, both the choice of mapper and parameters used severely affected the number of mapped reads (**Figure 1C**, see also **Table S1**). The most extreme case was observed with the Ψ-modified dataset, where *minimap2* was unable to map the majority of the reads (0–0.3% mapped reads) (**Figure 1C,D**, see also **Figure S1A**). By contrast, *GraphMap* ‘sensitive’ was able to map 35.5% of Ψ-modified base-called reads, with only a minor loss in accuracy (3%) (**Figure S1B**, see also **Table S2**), proving to be a more appropriate choice for highly modified datasets.

### Base-calling ‘error’ signatures can be used to predict RNA modification type

While base-calling ‘errors’ can be used to identify m<sup>6</sup>A RNA modified sites<sup>29,30,32</sup>, whether this approach is applicable for the detection of other RNA modifications, and whether these signatures could be employed to distinguish among distinct RNA modification types, is largely unknown. To this end, we systematically characterized the base-calling errors caused by the presence of m<sup>6</sup>A, Ψ, m<sup>5</sup>C and hm<sup>5</sup>C. We found that, regardless of the base-caller and mapper settings used, modified RNA sequences presented decreased quality scores (**Figure S1C–E**) and higher mismatch frequencies (**Figure 1E**), being these differences more prominent in Ψ-modified datasets. Principal component analysis of base-calling ‘errors’ of each modified dataset (m<sup>6</sup>A, Ψ, m<sup>5</sup>C and hm<sup>5</sup>C) -relative to

unmodified- showed that this difference was greatest in  $\Psi$ -modified datasets (**Figure 1F**), and maximized in datasets that were base-called with GU 3.0.3. Thus, we find that all four RNA modifications can be detected in direct RNA sequencing data; however, their detection is severely affected by the choice of both base-calling and mapping algorithms, and varies depending on the RNA modification type.

We then examined whether the base-called 'errors' observed in modified and unmodified datasets occurred in the modified position. We found that both  $m^6A$  and  $\Psi$  modifications led to increased mismatch frequencies at the modified site (**Figure 1G**), mainly in the form of U-to-C mismatches in the case of  $\Psi$  modifications (**Figure S1F**). By contrast,  $m^5C$  and  $hm^5C$  modifications did not appear in the form of increased mismatch frequencies at the modified site; rather, these modifications appeared in the form of increased mismatch frequencies in the neighbouring residues (position -1 and +1 in the case of  $m^5C$  modifications; position +1 in  $hm^5C$ ) (**Figure 1G**). Moreover, the base-called 'error' signatures of  $m^5C$  and  $hm^5C$  were also dependent on the sequence context (**Figure S1G**). Altogether, we found that all four RNA modifications studied ( $m^6A$ ,  $m^5C$ ,  $hm^5C$  and  $\Psi$ ) cause base-calling 'errors', and that these 'errors' follow specific patterns that depend on the RNA modification type.

#### **$\Psi$ modifications can be detected as U-to-C mismatches**

We then examined whether the results obtained using *in vitro* transcribed constructs would be applicable to *in vivo* RNA sequences. To this end, total RNA from *S. cerevisiae* was prepared for direct RNA sequencing (see *Methods*). Visual inspection of the mapped reads revealed a high proportion of base-calling errors present in 25s and 18s rRNAs, as could be expected from sequences that are highly enriched in RNA modifications (**Figure 2A**). By contrast, 5s and 5.8s rRNAs did not show such base-calling errors, in agreement with their low level of modification.

Then, we systematically analyzed base-called features (mismatch, deletion, insertion and per-base qualities) of rRNA modified sites relative to unmodified ones (**Figure 2B**), and found that all rRNA modification types consistently led to decreased per-base qualities at modified sites, suggesting that per-base qualities can be employed to identify RNA modifications, but not the underlying RNA modification type. We found that  $\Psi$  modifications caused significant variations in mismatch frequencies, in agreement with our observations using *in vitro* constructs. By contrast, other RNA modifications, such as 2'-O-methylcytidine (Cm) or 5-methylcytosine ( $m^5C$ ) did not appear in the form of increased mismatch frequencies at modified sites, but rather, in the form of increased insertions. In addition,  $\Psi$  modifications typically appeared in the form of U-to-C mismatches (**Figure 2C**, see also **Figure S2**), in agreement with our *in vitro* observations, whereas other RNA modifications such as 2'-O-methyladenosine (Am) did not cause mismatches with unique directionality. Thus, we conclude that distinct rRNA modification types can be detected in the form altered base-called features *in vivo*, and that their base-calling 'error' signature is dependent on the RNA modification type.

To confirm that the detected signal (U-to-C mismatches) in  $\Psi$  positions was caused by the presence of the  $\Psi$  modification, we compared rRNA modification profiles from wild type *S. cerevisiae* to those from snoRNA-knockout strains (snR3, snR34 and snR36) (**Figure 3A**, see also **Table S3**). Our results show that changes in rRNA modification profiles were consistently and exclusively observed in those  $\Psi$  positions reported as targets of each snoRNA. Moreover, the remaining  $\Psi$ -modified positions were not significantly altered by the lack of  $\Psi$  modifications guided by snR3, snR34 or snR36 (**Figure 3B**).

We then sequenced 3 additional *S. cerevisiae* strains depleted of snoRNAs (snR60, snR61 and snR62, respectively) guiding 2'-O-methylation (Nm) at specific positions (**Table S3**). In contrast to  $\Psi$  modifications, we found that 2'-O-methylations often caused increased errors not only at the modified position, but also at neighbouring positions (**Figure 3C**, see also **Figure S3A**). These errors disappeared in the knockout strain, confirming that neighbouring base-calling errors were indeed caused by the 2'-O-methylation (**Figure 3C**). On the other hand, while  $\Psi$  modifications mainly affected mismatch frequency, we observed that Nm modifications often affected several base-called 'error' features (mismatch, insertion and deletion frequency) (**Figure S3B**). Thus, we reasoned that combining all three features might improve the signal-to-noise ratio for the detection of 2'-O-methylated sites, and found that the combination of features led to improved detection of Nm-modified sites (**Figure 3D**).

#### **Current intensity variations cannot accurately predict the modified site**

We then wondered whether  $\Psi$  and Nm sites would also be detected at the level of current intensity changes. We observed that certain  $\Psi$  and Nm-modified sites, such as 25s: $\Psi$ 2129 or 25s:Am1133, showed drastic alterations in their current intensity values in the snoRNA-depleted strain, while no significant alteration was observed in control sites (**Figure 3E,F**). However, in other sites the distribution of current intensities did not significantly change in the knockout strain (18s: $\Psi$ 1187, **Figure 3E** lower panel) or did not differ in their mean (25s: $\Psi$ 2133, **Figure S4A**).

We hypothesized that deviations in current intensity alterations might not always be maximal in the modified site, but might sometimes appear in neighboring sites. To test this, we examined the difference in current intensity values along the rRNA molecules for each wild type-knockout pair (**Figure 4A**, see also **Figure S4B**). However, the highest deviations in current intensity were often not observed at the modified sites (**Figure 4A** lower panel). From all 6  $\Psi$  sites that were depleted in the 3 knockout strains studied, only 2 of them (25s: $\Psi$ 2826 and 25s: $\Psi$ 2880) showed a maximal deviation in current intensity in the modified site (**Figure 4B**, see also **Figure S4C**). Similarly, depletion of Nm sites led to changes in current intensity values, but the largest deviations were not observed at the modified site (**Figure S4C**). Thus, we conclude that current intensity-based methods can detect both  $\Psi$  and Nm RNA modifications; however, base-calling errors are a better choice to achieve single nucleotide resolution, at least in the case of  $\Psi$  RNA modifications.

#### **Detection of $\Psi$ and Nm modifications in individual reads**

Direct RNA sequencing produces current intensity measurements for each individual native RNA molecule. Thus, modification stoichiometries can be, in principle, estimated by identifying the proportion of reads with altered current intensity at a given site. To this end, we first examined the per-read current intensity values of wild type and knockout strains at the  $\Psi$ - and Nm-depleted sites. Despite the significant variability of current intensities across reads, we were able to observe robust differences in current intensities across strains at the depleted RNA modified sites at the per-read level (**Figure 4C**, upper panel). As a control, we performed the same analysis in  $\Psi$  sites unaffected by snoRNA depletion, finding no significant differences between wild type and knockout strains (**Figure 4C**, lower panel). However, in some sites such as 18s:1187, the per-read shifts in current intensity between the wild type and knockout strain were far more modest (**Figure S4D**). Principal Component Analysis (PCA) of the current intensity values of 15-mer regions that contained the modified site showed that the reads clustered into two distinct populations: the first population mainly comprised unmodified reads from the snoRNA-depleted strain, whereas the second comprised reads from the 3 other strains, which are mostly modified (**Figure 4C** right panels, see also **Figure S4E**).

To our surprise, we observed that *Nanopolish* software did not resquiggle the reads evenly across sites. For example, it failed at resquigging the majority of reads in the region surrounding 25s: $\Psi$ 2264 (**Figure S4D**). Thus, we examined whether *Tombo*, which uses global resquigging instead of local resquigging, might overcome this limitation. We found that *Tombo* resquigging led to a global increase in the proportion of resquigged reads (**Figure S5A**). Moreover, *Tombo* showed a uniform proportion of resquigged reads along the same transcript, whereas *Nanopolish* showed a variable proportion of resquigged reads depending on the site. Notably, *Tombo* was equally effective at resquigging both modified and unmodified reads, whereas *Nanopolish* preferentially resquigged unmodified reads relative to modified ones, biasing the unmodified:modified proportion up to 7:1 (**Figure S5B**). This uneven resquigging from *Nanopolish* implies that using *Nanopolish* for predicting RNA modification levels at individual sites may cause a dramatic bias in the predicted stoichiometry of individual sites. Thus, based on these results, we decided to adopt *Tombo* resquigging instead of *Nanopolish* resquigging for the prediction of RNA modification stoichiometries from individual RNA reads in all our downstream analyses.

### **Stoichiometry prediction using signal intensity, dwell time and trace**

$\Psi$  and Nm modifications can lead to significant alterations in the current intensity profiles at the modified region (e.g. 25s: $\Psi$ 2880, **Figure 4B-C**). However, in other sites such as 18s: $\Psi$ 1187, current intensity alone was insufficient to bin the reads into two separate clusters (**Figure S4D,E**), suggesting that, in addition to current intensity, other features might be needed to distinguish modified from unmodified reads.

Previous works predicting DNA modifications from individual nanopore reads typically relied on features such as signal intensity or dwell time to distinguish modified and unmodified read populations<sup>47–50</sup>. Here, in addition to these two features, we explored whether the use of ‘trace’ would improve our

ability to predict RNA modification stoichiometry. Trace (also termed 'base probability') represents the probability that a given signal intensity chunk may be originating from each of the 4 canonical bases (A, C, G and T/U). To this end, we first examined how the presence of  $\Psi$  and Nm modifications altered each of the features (signal intensity, dwell time and trace) in  $\Psi$  and Nm modified sites, both at snoRNA-targeted positions and control sites (**Figure S6**). We found that in addition to signal intensity, base probability (trace) was significantly different in all examined sites. Moreover, in some sites such as 25s: $\Psi$ 2264, trace was the most altered feature from those examined. By contrast, dwell time was not consistently different in snoRNA-targeted sites relative to wild type (e.g. 25s: $\Psi$ 2264, 25s: $\Psi$ 2826, 18s: $\Psi$ 1187).

We then proceeded to systematically benchmark the use of distinct features for RNA modification stoichiometry. To this end, we built *nanoRMS*, a software that extracts signal intensity, trace and dwell time from individual reads, and then predicts RNA modification stoichiometry by using distinct feature combinations as well as various machine learning algorithms. Firstly, we generated different mixes of modified (wild type) and unmodified (knockout) reads to simulate varying read stoichiometry (0, 20, 40, 60, 80 and 100%), for each of the  $\Psi$  and Nm positions for which knockouts were available (**Table S3**). Then, we examined how different supervised and unsupervised algorithms would predict the stoichiometry of each of the sites, and using distinct combinations of the 3 features (signal intensity, trace and dwell time) for each individual site (**Figure S5C**). Our results show that the combination of signal intensity and trace outperformed all the other feature combinations for predicting both  $\Psi$  and Nm modification stoichiometry, and that the supervised k-nearest neighbor (KNN) was the best performing algorithm. The k-means clustering algorithm (KMEANS) was the best-performing algorithm among the unsupervised clustering methods tested, although its performance in predicting  $\Psi$  modification stoichiometry was slightly better than in the case of Nm modification stoichiometry predictions. Overall, we find that *nanoRMS* can accurately predict  $\Psi$  and Nm RNA modification stoichiometry from individual RNA reads (**Figure 4D**), with predicted stoichiometry values that are similar to those that have been previously reported by Mass Spectrometry<sup>51</sup> (**Table S4**).

### **De novo prediction reveals a Pus4-dependent mitochondrial $\Psi$ rRNA modification**

The identification of RNA modification-specific signatures allows us to perform *de novo* prediction of  $\Psi$  RNA modifications transcriptome-wide using direct RNA sequencing. In this regard, *S. cerevisiae* mitochondrial rRNAs remains much less characterized than cytosolic rRNAs, with only 3 modified sites identified so far in *S. cerevisiae* LSU (21s)<sup>52</sup>, and none in SSU (15s) rRNAs. Thus, we hypothesized that direct RNA might reveal previously uncharacterized  $\Psi$ -modified sites in mitochondrial rRNAs. To this end, we first determined the 'error'-based thresholds (mismatch frequency and C mismatch frequency) that would distinguish unmodified uridines from pseudouridines in cytosolic rRNAs (**Figure 5A**). We then applied this filter to predict  $\Psi$  modifications on 15s rRNA and 21s rRNA, identifying two novel candidate  $\Psi$  sites (15s:854 and 15s:579) that displayed high modification frequency as well as U-to-C mismatch signature (**Figure 5B,C**).

To further confirm that the two predicted 15s rRNA sites are pseudouridylated, we developed nanoCMC-seq, a protocol that identifies  $\Psi$  modifications by coupling CMC probing with nanopore cDNA sequencing. This method allows capturing reverse-transcription drop-off information by sequencing only the first-strand cDNA molecules of CMC-probed RNAs using a customized direct cDNA sequencing protocol (**Figure 5D**, see also *Methods*). NanoCMC-seq captured known sites in cytoplasmic rRNA with a very high signal-to-noise ratio, as well as confirmed the existence of  $\Psi$  in position 854 and 579 of 15s rRNA, validating our *de novo* predictions using direct RNA sequencing (**Figure 5E**, see also **Figure S7A**).

We observed that 15s: $\Psi$ 854 was embedded in a similar sequence context and structure as the t-arm of tRNAs, which contains a pseudouridylated ( $\Psi$ 55) position placed by Pus4 (**Figure 5F**). Given the resemblance between these two sequences and structures, we hypothesized that Pus4 might be responsible for this modification. To validate our hypothesis, we sequenced total RNA from a Pus4-deficient *S. cerevisiae* strain, finding that the 15s:854 position loses its mismatch signature upon deleting Pus4 gene, confirming that this site is not only pseudouridylated, but also that it is Pus4-dependent (**Figure 5G**, see also **Figure S7B**). Additionally, we observed that previously reported Pus4 target sites (TEF1:239,TEF2:239)<sup>3-5</sup> completely lost their mismatch signature in Pus4 knockout cells (**Figure S7B,C**), confirming that our method is able to capture previously reported Pus4-dependent  $\Psi$  sites, in addition to previously unknown ones.

### **rRNA modification profiles do not vary upon oxidative or thermal stress**

Ribosomal RNAs are extensively modified as part of their normal maturation, and their modification landscape is relatively well-defined for a series of organisms<sup>39,53-57</sup>. Despite the central role that rRNA molecules play in protein translation, recent evidence has shown that rRNA modifications are in fact dynamically regulated<sup>58,59</sup>, and that their alterations can lead to disease states<sup>41,42,60-66</sup>. Moreover, the stoichiometry of some pseudouridylated and 2'-O-methylated rRNA sites is cell-type dependent, suggesting that rRNA modifications may be an important source of ribosomal heterogeneity<sup>43,51,54,67-69</sup>. However, a systematic and comprehensive analysis of which environmental cues may lead to changes in rRNA modification stoichiometries, which RNA modifications may be subject to this tuning, and to which extent, is largely missing.

To assess whether rRNA modification profiles change in response to environmental stimuli, we treated *S. cerevisiae* cells with diverse environmental cues (oxidative, cold and heat stress) and sequenced their RNA using direct RNA sequencing. Firstly, we confirmed that the rRNA modification profiles from independent biological replicates were highly reproducible (Pearson  $r^2=0.976-0.996$ , see also **Figure S8**). Then, we examined whether exposure to stress would lead to significant changes in base-calling 'errors' in rRNA molecules, finding no significant differences in rRNA modification profiles between normal and stress conditions (**Figure 6A**). In contrast, we recapitulated previously reported changes in snRNA  $\Psi$  modifications upon exposure to environmental cues<sup>4</sup> (**Figure 6B**, see also **Figure S7D**), as well as identified 8 additional  $\Psi$  modification sites in snRNAs and snoRNAs whose

stoichiometry varies upon heat exposure, which had not been previously described (**Figure 6**, see also **Figure S7E and Table S5**)<sup>3,4,38,70</sup>. Overall, our approach confirmed previous reports and predicted novel  $\Psi$  sites in ncRNAs whose modification levels vary upon heat shock exposure (**Figure 6B-D**, see also **S7D-E**), but did not identify any rRNA modified site to be varying in its stoichiometry upon any of the tested stress conditions.

#### **rRNA modification profiles do not vary across translational repertoires**

Next, we questioned whether pseudouridylation changes in distinct translational repertoires may be more nuanced, in that  $\Psi$  levels may differ between rRNAs present in different translational fractions along a polysome gradient, which would not be detected when examining rRNAs as a whole. To test this, we sequenced both total (input) and polysomal rRNAs from untreated and H<sub>2</sub>O<sub>2</sub>-treated yeast cells (**Figure S7F**). However, we observed no significant changes in  $\Psi$  rRNA modification profiles when comparing rRNAs from actively translating ribosomes in untreated versus H<sub>2</sub>O<sub>2</sub>-treated cells (**Figure S7G**).

In an attempt to further dissect the different translational repertoires into a higher number of rRNA pools, we sequenced: i) rRNAs from unassembled free rRNA fractions (F1), ii) rRNAs from 40s and 60s subunits (F2), iii) rRNAs from monosomal fractions (F3) and iv) rRNAs from polysomal fractions (F4) (**Figure 6E**). While two positions showed slightly decreased levels of  $\Psi$  (5.8s: $\Psi$ 73 and 25s: $\Psi$ 776) in the free rRNA fraction (F1) compared to assembled ribosomes and/or subunits, no significant changes were observed across the other translational fractions (**Figure 6F**, see also **Figure S7H**). Globally, these results indicate that differential rRNA modifications are likely not the mechanism employed by yeast cells to adapt to environmental stress conditions, in agreement with previous observations<sup>3</sup>.

#### **De novo prediction of $\Psi$ modifications in mRNAs**

Ribosomal RNAs are modified at very high stoichiometries<sup>51,54</sup>. By contrast, other molecules such as mRNAs are modified at lower stoichiometries, making the detection of their RNA modifications a much more challenging task<sup>21</sup>. To ascertain whether our methodology would be applicable to lowly modified RNA sites, such as those present in mRNAs, we first assessed the performance of *nanoRMS* in RNA molecules that contained  $\Psi$  RNA modifications at low RNA modification stoichiometries (0, 3, 7 and 20%) (**Figure 7A**, see also *Methods*). The relative incorporation of  $\Psi$  RNA modifications was validated using Mass Spectrometry. We then examined the quantitative performance of *nanoRMS* under low stoichiometry conditions using both KNN and k-means, finding that the combination of signal intensity and trace features yielded the most accurate results in terms of stoichiometry prediction (**Figure 7B**), in agreement with our previous results (**Figure S5C**).

Next, we sequenced polyA(+)-selected RNA from *S. cerevisiae* wild type, Pus1 knockout, Pus4 knockout and heat stress-exposed strains using direct RNA sequencing, in biological duplicates. Considering that mRNA sites are lowly modified, we restricted our *de novo* identification of mRNA  $\Psi$

397 sites to those whose base-calling 'error' features significantly changed between pairwise conditions  
398 (**Figure 7C**, see also *Methods*), met the pseudouridine 'error' signature, and had a minimum coverage  
399 of 30 reads in both conditions and biological replicates (**Table S6**, see also *Methods*). Through this  
400 approach, we predicted 13 Pus1-dependent  $\Psi$  mRNA modifications, 14 Pus4-dependent  $\Psi$  mRNA  
401 modifications, 17 heat stress-dependent  $\Psi$  mRNA modifications and 16 heat stress-dependent  $\Psi$   
402 ncRNA modifications, respectively (**Figure 7D-G** left panels, see also **Tables S7-10**), some of which  
403 were not previously reported to be  $\Psi$ -modified.

404  
405 *NanoRMS* recovered 11% of previously reported Pus1-dependent  $\Psi$  sites as well as 75% Pus4-  
406 dependent  $\Psi$  sites, in addition to predicting 10 not previously reported Pus1 and 11 not previously  
407 reported Pus4-dependent mRNA  $\Psi$ -modified sites (**Table S7** and **S8**). These novel predicted  $\Psi$   
408 mRNA sites displayed similar mismatch signatures to those observed in previously reported  $\Psi$  sites  
409 (**Figure 7D-E**, right panels), were highly replicable across biological replicates, and their signature  
410 disappeared in Pus1 or Pus4 knockout strains. Similarly, *nanoRMS* was able to capture previously  
411 reported heat-responsive  $\Psi$  sites present in mRNAs and ncRNAs, which resulted in predicting 17  
412 heat-responsive  $\Psi$  mRNAs sites, among which 6 of them were previously reported  $\Psi$  sites (**Figure**  
413 **7F**, see also **Table S9**), as well as 16 heat-responsive  $\Psi$  ncRNAs sites, from which 10 were  
414 previously reported  $\Psi$  sites (**Figure 7G**, see also **Table S10**).

415  
416 Surprised by the relatively poor overlap between our predictions and previously reported Pus1 mRNA  
417  $\Psi$ -modified sites (3 out of 16 sites), as well as between predicted and previously reported heat stress-  
418 dependent sites (7 out of 128 sites), we inspected the individual per-read features at previously  
419 reported Pus1- and heat stress-dependent sites (**Figure S9A,B**). Indeed, the  $\Psi$  sites that *nanoRMS*  
420 did not report as Pus1 or heat stress-dependent were not significantly different for any of the features  
421 examined (current intensity, dwell time or trace). Thus, we wondered whether some of these sites  
422 might have been misassigned as Pus1 or heat stress-dependent by previous works. A closer  
423 examination of the overlap between  $\Psi$  sites predicted by the two previously published studies using  
424 CMC probing coupled to Illumina sequencing<sup>3,4</sup>, which we used to define the set of 'previously  
425 reported Pus1-, Pus4- and heat stress-dependent  $\Psi$  sites', showed that the overlap was in fact very  
426 poor (**Figure S9C**), both when examining the set of predicted mRNA and ncRNA  $\Psi$  sites (7% and  
427 17%, respectively), as well as when examining the sets of predicted Pus1- and Pus4-dependent  
428 mRNA and ncRNA  $\Psi$  sites (6% and 50%, respectively). Altogether, our approach detected 100% of  
429 Pus1- and Pus4-dependent sites that were identified by both studies, but very few of those that were  
430 identified by only one of the studies. Thus, we conclude that the poor overlap between our results and  
431 previously reported  $\Psi$  sites is in fact a direct consequence of the poor overlap between the set of  
432 predicted Pus1-, Pus4- and heat stress-dependent mRNA and ncRNA  $\Psi$  sites by the two previous  
433 studies (**Figure S9C**).

434  
435 Finally, we applied *nanoRMS* to predict the modification stoichiometry of all  $\Psi$  sites predicted in  
436 mRNAs and ncRNAs. Reads were classified based on the per-read signal intensity and trace features

from positions -1, 0, and +1 using the k-means unsupervised clustering algorithm (**Figure 7H-K**). As expected, per-read stoichiometry predictions were low in non-targeted  $\Psi$  sites. By contrast, predicted  $\Psi$  Pus1/Pus4/heat stress-dependent sites (which included all  $\Psi$  sites) typically showed significant RNA modification stoichiometry changes, ranging from 5 to 50% change in their  $\Psi$  modification stoichiometries between the two conditions.

Altogether, we find that differential 'error'  $\Psi$  signatures are a useful approach to identify dynamic  $\Psi$  RNA modifications across two conditions even at low stoichiometry sites, and that *nanoRMS* can be used to *de novo* predict and quantify the RNA modification stoichiometry dynamics, both in previously reported  $\Psi$  sites as well as in *de novo* predicted  $\Psi$  sites.

## DISCUSSION

RNA modifications regulate a wide range of biological processes<sup>71-73</sup>. They can modulate the fate of RNA molecules by altering mRNA splicing<sup>74-76</sup> or mRNA decay<sup>77,78</sup>, as well as affect major cell and organism-level decisions, such as cellular differentiation<sup>79,80</sup> and sex determination<sup>18,81,82</sup>. While the biological relevance of RNA modifications is out of question, a major difficulty in studying them has been the need for tailored protocols to map each modification individually<sup>20,83</sup>. In this context, direct RNA nanopore sequencing can overcome many of the limitations that NGS-based methods suffer from, as it can sequence full-length native RNA molecules, including their RNA modifications.

Direct RNA nanopore sequencing has been successfully applied in a wide variety of organisms<sup>29-31,84-87</sup>. However, the detection of distinct RNA modification types in individual native RNA molecules is still an unsolved challenge. While both current intensity-based and 'error'-based methods have proven useful strategies to detect RNA modifications, these have been mainly focused on the detection of m<sup>6</sup>A<sup>29-31,33</sup>-, and are typically unable to predict which RNA modification type they are in fact detecting (e.g. m<sup>6</sup>A,  $\Psi$ , Am or m<sup>5</sup>C)<sup>28,50</sup>. Moreover, current algorithms to study RNA modifications using direct RNA sequencing are not quantitative.

To overcome these limitations, here we first explored how distinct RNA modifications may affect direct RNA nanopore signals and base-calling 'errors'. We find that different RNA modification types (e.g.  $\Psi$  versus m<sup>5</sup>C) produce distinct yet characteristic base-calling 'error' signatures, both *in vitro* (**Figure 1, S1F**) and *in vivo* (**Figure 2**). Consequently, base-calling errors can be used not only to predict whether a given site is modified or not, but also to identify the underlying RNA modification type. While base-calling signatures depend to some extent on the surrounding sequence context, we find that  $\Psi$  modifications lead to robust U-to-C mismatch signatures, which can be exploited for *de novo* prediction of  $\Psi$  modifications (**Figure 5**). Through this approach, we identified two previously unreported  $\Psi$  modifications in yeast 15s mitochondrial rRNA (15s:579 and 15s: $\Psi$ 854), as well as confirmed reported  $\Psi$ -modified sites in rRNAs, snRNAs and mRNAs (**Figures 3-7**). Moreover, we revealed that Pus4, which was previously thought to modify only tRNAs and mRNAs, is the enzyme

responsible for placing  $\Psi$ 854 in mitochondrial rRNA. These findings were further validated using nanoCMC-seq, a novel orthogonal method that can detect  $\Psi$  modifications with single nucleotide resolution by coupling CMC probing to nanopore cDNA sequencing (**Figure 5D**).

While  $\Psi$  modifications can be detected both in the form of base-calling 'errors' and altered current intensities (**Figures 3-4**), we observe that the latter does not provide single nucleotide resolution, with maximal current intensity shifts often seen a few nucleotides away from the real modified site. Thus, current intensity-based methods alone may suffer from imprecisions in the assignment of the RNA-modified site. Here we propose that the combination of both approaches is the optimal design to obtain stoichiometric information of  $\Psi$ -modified sites with single nucleotide resolution. Specifically, we show that once the site has been located using base-calling error features, per-read features (current intensity and trace) from the regions surrounding  $\Psi$  or Nm-modified site are sufficient to robustly bin the reads into two separate clusters (modified and unmodified), and provide good estimates of  $\Psi$  and Nm modification stoichiometries (**Figure 4D** and **7B**).

One surprising feature of base-calling 'errors' is that fully modified sites do not always lead to same mismatch frequencies, suggesting that mismatch frequencies alone cannot be used per se as an estimation of the stoichiometry of the site (**Figure 2B**). While within the same sequence context, higher mismatch frequencies correspond to higher modification levels, this same rule cannot be used to compare across distinct RNA-modified sites. We speculate that the differences observed in mismatch frequency across different sites might be in fact a consequence of the distinct deviations in current intensity of the modified k-mer relative to unmodified counterparts (**Figure S9D**).

Finally, we should note that while *nanoRMS* allows predicting and studying the dynamics of diverse RNA modifications in a quantitative manner, there are caveats and limitations, leaving ample room for future improvements. First, not all RNA modifications lead to strong alterations in the base-calling features and/or current intensity patterns, such as 2'-O-methylcytosine (Cm), which is poorly detected in direct RNA sequencing datasets, compared to other RNA modifications (**Figure 2C**). Second, the detection of RNA modifications is partly dependent on the sequence context; for example, we were unable to detect 25s:Gm908 (**Figure S3**). Similarly, some  $\Psi$ -modified sites, such as 18s: $\Psi$ 1187, cause weaker alterations in base-calling features and current intensity shifts than other  $\Psi$ -modified positions (**Figures 3-4**), although this limitation can be alleviated by the incorporation of additional features into the model (**Figure S5C**). Third, not all RNA modifications lead to base-calling errors with single nucleotide resolution, as with pseudouridine. For example, 2'-O-methylations often affect neighboring bases (**Figure 3C** and **S4A**), making it challenging to *de novo* predict modified sites without any prior information. Fourth, stoichiometry prediction is heavily affected by the choice of resquigging algorithms (**Figure S5** and **S10**). For example, we were unable to predict stoichiometry in 25s: $\Psi$ 2264 when using resquigging due to the low number of reads that the *Nanopolish* algorithm was able to resquiggle (**Figure S4E**); however, this limitation could be overcome when using *Tombo* resquigging, leading to stoichiometry predictions similar to those observed using Mass Spectrometry

(Figure 4D). Finally, we should note that while *nanoRMS* was successful at detecting RNA modification stoichiometry changes as low as 5-10% (Figure 7), the detection of RNA modification changes in low modification stoichiometry sites was only possible when using pairwise comparisons.

Despite these challenges and limitations, our work provides a framework for the systematic and comprehensive analysis of the epitranscriptome with single molecule resolution, showing that direct RNA sequencing can be employed to estimate  $\Psi$  and Nm modification stoichiometry as well as to *de novo* predict  $\Psi$  RNA modifications transcriptome-wide, in rRNAs, ncRNAs and mRNAs. Future work will be needed to functionally dissect the biological roles and dynamics of RNA modifications, to better comprehend how and when the epitranscriptome is tuned to regulate diverse cellular functions.

## FIGURE LEGENDS

**Figure 1. Systematic analysis of base-calling and mapping algorithms for the detection of RNA modifications in direct RNA sequencing datasets.** (A) Overview of the synthetic constructs used to benchmark the algorithms, which included both unmodified (UNM and UNM-S) and modified ( $m^6A$ ,  $m^5C$ ,  $hm^5C$  and  $\Psi$ ) sequences. For each dataset, we performed: i) comparison of base-calling algorithms, ii) comparison of mapping algorithms, iii) detection of RNA modifications using base-called features and iv) comparative analysis of features to distinguish similar RNA modifications. (B) Barplots comparing the percentage of base-called reads using 4 different base-calling algorithms in 6 different unmodified and modified datasets. (C) Relative proportion of base-called and mapped reads using all possible combinations (16) of base-callers and mappers included in this study, for each of the 6 datasets analyzed. (D) IGV snapshots illustrating the differences in mapping for 3 distinct datasets: UNM,  $m^6A$ -modified and  $\Psi$ -modified when base-called with GU 3.0.3. Positions with mismatch frequencies greater than 0.1 have been colored, gray represents match to reference. (E) Comparison of global mismatch frequencies using different base-calling algorithms, for the 6 datasets analyzed. Box, first to last quartiles; whiskers, 1.5x interquartile range; center line, median; points, outliers; violin, distribution of density. (F) Principal Component Analysis (PCA) using as input the base-calling error features of quality, mismatch frequency and deletion frequency in positions -2, -1, 0, 1 and 2, for all datasets base-called with GU 3.0.3 and AL 2.1.7 and mapped with GraphMap and minimap2 on sensitive settings. Only k-mers that contained a modification at position 0, and no other modifications in the 5-mer, were included in the analysis, and the equivalent set of unmodified k-mers was used as a control. (G) Mismatch frequency of each position of the 5-mers centered in the modified position (position 0). Box, first to last quartiles; whiskers, 1.5x interquartile range; center line, median; points, outliers. See also Figure S1.

**Figure 2. RNA modifications can be detected in yeast ribosomal RNA in the form of base-calling errors, and each RNA modification type shows a distinct 'error' signature.** (A) IGV snapshots of yeast ribosomal subunits 5s, 5.8s, 18s and 25s. Known modification sites are indicated below each snapshot and nucleotides with mismatch frequencies greater than >0.1 have been

colored and gray represents match to reference or no mismatch **(B)** Comparison of base-calling features (base quality, mismatch, deletion and insertion frequency) from distinct RNA modification types present in yeast ribosomal RNA. The most descriptive base-calling error per modification is outlined in red. Only RNA modification sites without additional neighboring RNA modifications in the 5-mer were included in the analysis:  $\Psi$  (n=37), Gm (n=8),  $m^1A$  (n=2), Am (n=14),  $m^5C$  (n=2), Cm (n=8),  $ac^4C$  (n=2), Um (n=7). Box, first to last quartiles; whiskers, 1.5x interquartile range; center line, median; dots: individual data points. **(C)** Ternary plots and barplots depicting the mismatch directionality for selected rRNA modifications ( $\Psi$ , Am, Cm, Gm).  $\Psi$  rRNA modifications tend towards U-to-C mismatches while Am, Cm and Gm modifications did not show specific mismatch directionality patterns. See also Figure S2 and S3.

**Figure 3. Pseudouridylation and 2'-O-methylations cause systematic base-calling 'errors' as well as altered current intensities, and their signature disappears upon depletion of snoRNAs guiding the modification.** **(A)** IGV snapshots of wild type and three snoRNA-depleted strains depicting the site-specific loss of base-called errors at known  $\Psi$  target positions (indicated by asterisks). Nucleotides with mismatch frequencies greater than 0.1 have been colored. **(B)** Comparison of snoRNA knockout mismatch frequencies for each base, relative to wild type, with snoRNA targets sites indicated in red, and non-target sites in gray. **(C)** IGV snapshots of wild type and three snoRNA knockout yeast strains depicting the site-specific loss of base-calling errors at known Nm target positions. Nucleotides with mismatch frequencies greater than 0.1 have been colored. **(D)** Comparison of snoRNA knockout summed error frequencies for each base, relative to wild type, with snoRNA targets sites indicated in red, neighboring sites in blue and non-target sites in gray. **(E,F)** Distributions of per-read current intensity at known  $\Psi$ -modified (E), 2'-O-methylated (F) and negative control sites. Current intensities at  $\Psi$  and 2'-O-methylated positions were altered upon deletion of specific snoRNAs relative to wild type, whereas no shift was observed in control sites.

**Figure 4. Loss of specific  $\Psi$  rRNA modifications causes deviations in current intensity in regions surrounding the  $\Psi$  sites.** **(A)** Current intensity changes along the 25s rRNA molecule upon snR3 depletion, relative to the wild type strain. In the lower panel, a zoomed subset focusing on the two regions with the most significant current intensity deviations is shown; the first one comprising the 25s: $\Psi$ 2129 and 25s: $\Psi$ 2133 sites, and the second one comprising the 25s: $\Psi$ 2264 site. **(B)** Comparison of current intensities in the 15-mer regions surrounding  $\Psi$  and 2'-O-methyl knockout sites, for each of the 4 strains. The dotted vertical line indicates the modified position. See also Figure S4 for current intensity changes in other knockout strains and sites. **(C)** Per-read current intensity analysis centered at the 25s: $\Psi$ 2880 site targeted by snR34 (upper panel) and a control site, 25s: $\Psi$ 2880, which is not targeted by any of the knockouts (lower panel). For each site, Principal Component Analysis was performed using 15-mer current intensity values, and the corresponding scatterplot of the two first principal components (PC1 and PC2) is shown on the right, using as input the same read populations as in the left panels. Each dot corresponds to a different read, and is colored according to the strain. **(D)** Predicted stoichiometry of  $\Psi$ - and Nm-modified sites using a k-nearest neighbors (KNN) algorithm

trained to classify the reads into 2 classes: modified or unmodified. The features used to predict modifications status of every read from which stoichiometry was calculated were signal intensity (positions -1,0,+1) and trace (positions -1,0,+1). See also Figures S4 and S5.

**Figure 5. De novo prediction of  $\Psi$  modifications reveals a novel Pus4-dependent mitochondrial rRNA modification.**

**(A)** Density distributions of mismatch and C mismatch frequency in unmodified uridine (red) and pseudouridine (cyan) positions. The dashed lines represent the optimal cutpoints between two groups determined by maximizing the Youden-Index. In the right panel, the ROC curve illustrates the sensitivity and specificity at these two cutpoints. **(B)** IGV coverage tracks of the 15s mitochondrial rRNA, including a zoomed version showing the tracks centered at the 15s:854 and 15s:579 sites, in two biological replicates. Nucleotides with mismatch frequencies greater than 0.15 have been colored. **(C)** Location of the putative  $\Psi$ 854 modified site in the yeast mitochondrial ribosome. The PDB structure shown corresponds to 5MRC. **(D)** Validation of putative  $\Psi$  sites with nanoCMC-Seq, which combines CMC treatment with Nanopore cDNA sequencing in order to capture RT-drops that occur at  $\Psi$ -modified sites upon CMC probing. RT-drops are defined by counting the number of reads ending (3') at a given position. **(E)** Predicted  $\Psi$  sites U854 and U579 (orange) in the 15s rRNA are validated using nanoCMC-seq (upper panel). Dashed lines indicate the CMC-score threshold used for determining the positive sites (upper panel). As a control, we analysed the nanoCMC-seq results in other rRNAs (lower panel), finding that all positions with a significant CMC Score (>25) correspond to known  $\Psi$  rRNA modification sites (blue). See also Figure S7A for CMC scores in additional rRNA transcripts. **(F)** The candidate  $\Psi$ 854 site is located at the 852-860 loop of the 15s rRNA, which resembles the t-arm of the tRNAs that is modified by Pus4. The binding motif of Pus4 (RRUUCNA) matches the motif surrounding the 854U site<sup>4</sup>. **(G)** Scatterplot of mismatch frequencies in WT and Pus4KO cells, showing that the only significant position affected by the knockout of Pus4 is 15s:U854 (left panel). IGV coverage tracks showing that Pus4 knockout leads to depletion of the mismatch signature in the 15s:854 position (right panel), but not at the 15s:579 position.

**Figure 6. Comparative analysis of yeast rRNA and snRNA  $\Psi$  modifications upon distinct environmental stresses identifies known and previously unknown heat-sensitive snRNA and snoRNA  $\Psi$  modifications.**

**(A)** Comparison of mismatch frequencies for all rRNA bases from untreated or yeast exposed to oxidative stress (H<sub>2</sub>O<sub>2</sub>, left panel), cold stress (4°C, middle panel) or heat stress (45°C, right panel). Each dot represents a uridine base. All rRNA bases from cytosolic rRNAs were included in the analyses. **(B)** Comparison of mismatch frequencies in untreated versus stressed-exposed yeast cells (oxidative, cold or heat), in previously reported ncRNA  $\Psi$  sites<sup>3,4</sup>. **(C)** Stress scores in sn/snoRNA  $\Psi$  sites calculated by  $\Delta$  mismatch frequency between heat shock and WT. **(D)** IGV snapshots of normal condition (rep1 and rep2) and heat shock condition (rep1 and rep2) yeast cells zoomed into the known sn/snoRNA  $\Psi$  positions (indicated by an asterisk). Nucleotides with mismatch frequencies greater than 0.1 have been colored. Coverage for each position/condition is given on the top left of each row. **(E)** Profiles of ribosomal fractions isolated from yeast grown under

normal conditions, using sucrose gradient fractionation, including free rRNAs which are not assembled into ribosomal subunits (F1), rRNAs from 40s and 60s subunits (F2), rRNAs extracted from monosomal fractions (F3) and polysome fractions (F4). **(F)** IGV snapshots of the two  $\Psi$  sites that change stoichiometry between translational fractions and four representative  $\Psi$  sites that show no significant change. Nucleotides with mismatch frequencies greater than 0.1 have been colored. See also Figure S7 and S8.

**Figure 7. Quantitative prediction of pseudouridine stoichiometry transcriptome-wide and systematic benchmarking of *nanoRMS* using RNA molecules with diverse modification stoichiometries.** **(A)** LC-MS/MS validation of pseudouridine incorporation at different proportions (0%, 3%, 20%, 100%) in the *in vitro* transcribed products, relative to the expected incorporation ( $\Psi$ TP relative to UTP) (left panel). Dotplot illustrates the mismatch frequency distribution of the uridine positions in the *in vitro* transcribed products incorporated with different concentrations of  $\Psi$  (right panel). Each dot represents one uridine position. **(B)** Stoichiometry predictions of the  $\Psi$  incorporated *in vitro* transcription products using two different algorithms (KNN and k-means) with different current information (middle right and right panels). **(C)** Conditions and strains used to predict  $\Psi$  mRNA modifications transcriptome-wide. **(D-K)** Transcriptome-wide  $\Psi$  RNA modification predictions and predicted stoichiometries in mRNAs and ncRNAs, for Pus1-dependent mRNA  $\Psi$  sites (D,H), Pus4-dependent mRNA  $\Psi$  sites (E,I), heat stress-dependent mRNA  $\Psi$  sites (F,J) and heat stress-dependent ncRNA  $\Psi$  sites (G,K). **(D-G)** Venn diagrams depict the overlap between  $\Psi$  sites predicted by our analysis and the previously reported pseudouridine sites. IGV snapshots of reported and not previously reported predicted sites illustrate the absence of the mismatch signature in the Pus1 (D) or Pus4 (E) knockout samples as well as under normal conditions, relative to heat stress conditions in mRNA (F) and ncRNA (G). The reported or predicted  $\Psi$  site is indicated by an asterisk. Nucleotides with mismatch frequencies greater than 0.15 have been colored. We should note that IGV snapshots that show a reference “A” with mismatch signature to G are genes that are in the minus strand (and thus are in reality positions showing U-to-C mismatch signatures). **(H-K)** Quantitative analysis of previously reported and *de novo* predicted  $\Psi$  sites in mRNAs and ncRNAs. In the left panels, comparative scatterplots of mismatch frequency illustrate differentially modified sites of reported and *de novo* predicted  $\Psi$  sites. In the right panels, stoichiometry prediction differences between WT and knockout strains (H-I) or between normal and heat stress conditions (J-K) are shown as boxplots. Box, first to last quartiles; whiskers, 1.5x interquartile range; center line, median; points, individual  $\Psi$  sites. See also Figure S9.

## METHODS

### Yeast culturing

*Saccharomyces cerevisiae* (strain BY4741) was grown at 30°C in standard YPD medium (1% yeast extract, 2% Bacto Peptone and 2% dextrose). The deletion strains snR3Δ, snR34Δ and snR36Δ were generated on the background of the BY4741 strain by replacing the genomic snoRNA sequence with

a *kanMX4* cassette as detailed in Parker et al.<sup>88</sup>. Cells were then quickly transferred into 50 mL pre-chilled falcon tubes, and centrifuged for 5 minutes at 3,000 g in a 4°C pre-chilled centrifuge. Supernatant was discarded, and cells were flash frozen. For thermal stress, *Saccharomyces cerevisiae* BY4741 cultures were grown in 4 mL of YPD overnight at 30°C. The next day, cultures were diluted to 0.0001 OD600 in 200 mL of YPD and grown overnight at 30°C shaking (250 rpm). When the cultures reached an OD600 of 0.4-0.5, the cultures were divided into 3 x 50 mL subcultures, which were then incubated at 30°C (control), 45°C (heat shock) or 4°C (cold shock) for 1 hour. Cells were collected by pelleting and snap freezing. For the analysis of rRNAs modifications across polysomal fractions, yeast BY4741 starter cultures were grown in 6 mL YPD medium at 30°C with shaking (250 rpm) overnight. 100 mL of fresh YPD medium was inoculated with 10 µL of the stationary culture in a 250 mL erlenmeyer flask, in biological duplicates. Cells were incubated at 30°C with shaking (250 rpm) until the cultures reached mid-exponential growth phase ( $O.D_{600} \sim 0.4-0.6$ ). Yeast cells were then treated with 1 mM H<sub>2</sub>O<sub>2</sub> or left without treatment (control) for 30 minutes. 1 mL of cycloheximide stock solution (10 mg/mL) was added to each culture. Pus4 knockout strains (BY4741 MATa *pus4::KAN*) and its parental strain were obtained from the Yeast Knockout Collection (Dharmacon) and grown under standard conditions in YPD (1% [w/v] yeast extract, 2% [w/v] peptone supplemented with 2% glucose) at 30°C unless stated otherwise.

#### **Total RNA extraction from yeast cultures**

*Saccharomyces cerevisiae* BY4741 cells (strains: snR3Δ, snR34Δ snR36Δ, snR60Δ, snR61Δ, snR62Δ and WT) were harvested via centrifugation at 3000 rpm for 1 minute, followed by two washes with water. RNA was purified from pelleted cells using a MasterPure Yeast RNA extraction kit (Lucigen, MPY03100), according to manufacturer's instructions. Total RNA was then treated with Turbo DNase (Thermo, #AM2238) with a subsequent RNAClean XP bead cleanup prior to starting the library preparation. For stress conditions and the Pus4KO strain, flash frozen pellets were resuspended in 700 µL Trizol with 350 µL acid washed and autoclaved glass beads (425-600 µm, Sigma G8772). The cells were disrupted using a vortex on top speed for 7 cycles of 15 seconds (the samples were chilled on ice for 30 seconds between cycles). Afterwards, the samples were incubated at room temperature for 5 minutes and 200 µL chloroform was added. After briefly vortexing the suspension, the samples were incubated for 5 minutes at room temperature. Then they were centrifuged at 14,000 g for 15 minutes at 4°C and the upper aqueous phase was transferred to a new tube. RNA was precipitated with 2X volume Molecular Grade Absolute ethanol and 0.1X volume Sodium Acetate. The samples were then incubated for 1 hour at -20°C and centrifuged at 14,000 g for 15 minutes at 4°C. The pellet was then washed with 70% ethanol and resuspended with nuclease-free water after air drying for 5 minutes on the benchtop. Purity of the total RNA was measured with the NanoDrop 2000 Spectrophotometer. Total RNA was then treated with Turbo DNase (Thermo, #AM2238) with a subsequent RNAClean XP bead cleanup.

## **mRNA extraction from yeast cultures**

*Saccharomyces cerevisiae* BY4741 (strains: BY4741 MATa *pus4::KAN*, BY4741 MATa *pus1::KAN* and BY4741 MATa) were cultured up to log phase at 30°C. The cultures were then divided into two flasks and cultivated at 30°C or 45°C for 1 hour. The cells were harvested via centrifugation at 3,000 rpm for 5 minutes and snap frozen. Total RNA was purified from pelleted cells using a MasterPure Yeast RNA extraction kit (Lucigen, MPY03100), according to manufacturer's instructions. Total RNA was then DNase-treated (Ambion, AM2239) at 37°C for 20 minutes with a subsequent clean up using RNeasy MinElute Cleanup Kit (Qiagen, 74204). 70-100 ug of total RNA was subjected to double polyA-selection using Dynabeads Oligo(dT)25 (Invitrogen, 61002) and finally eluted in ice-cold 10 mM Tris pH 7.5.

## **Polysome gradient fractionation and rRNA extraction**

Yeast pellets from 100 mL cultures were washed with 6 mL of ice-cold Polysome Extraction Buffer (PEB), which contained 20 mM Tris-HCl pH 7.4, 100 mM KCl, 10 mM MgCl<sub>2</sub>, 0.5 mM DTT, 0.1 mg/mL cycloheximide and 100 U/mL RNase inhibitors (RNaseOUT, Invitrogen, #18080051). Cells were centrifuged for 5 minutes at 3,000 g at 4°C. Washing was repeated by adding 6 mL of ice-cold PEB, followed by centrifugation. Cells were then resuspended in 700 µL of ice-cold PEB, and transferred into pre-chilled 2 mL Eppendorf tubes containing 450 µL of pre-chilled RNase-free 425-600 µm diameter glass beads (Sigma G8772). Cells were lysed by vortexing at maximum speed for 5 minutes at 4°C, followed by centrifugation also at maximum speed at bench centrifuge for 5 minutes at 4°C. 10% of the supernatant was aliquoted into Trizol for total RNA isolation, and kept at -80°C, which was later used as input. The remaining volume, corresponding approximately to  $8 \times 10^8$  cells, was subsequently loaded onto the sucrose gradient. Linear sucrose gradients of 10-50% were prepared using the Gradient Station (BioComp). Briefly, SW41 centrifugation tubes (Beckman, Ultra-Clear™ 344059) were filled with Gradient Solution 1 (GS1), which consisted of 20 mM Tris-HCl pH 7.4, 100 mM KCl, 10 mM MgCl<sub>2</sub>, 0.5 mM DTT, 0.1 mg/mL cycloheximide and 10% w/v RNase-free sucrose. Solutions GS1 and GS2 were prepared with RNase-DNase free UltraPure water and filtered with a 0.22 µm filter. The tube was then filled with 6.3 mL of Gradient Solution 2 (GS2) layered at the bottom of the tube, which consisted of 20 mM Tris-HCl pH 7.4, 100 mM KCl, 10 mM MgCl<sub>2</sub>, 0.5 mM DTT, 0.1 mg/mL cycloheximide and 50% w/v RNase-free sucrose. The linear gradient was formed using the tilted methodology, with the Gradient Station Maker (Biocomp). Once the gradients were formed, 350 µL of each lysate was carefully loaded on top of the gradients, and tubes were balanced in pairs, placed into pre-chilled SW41Ti buckets and centrifuged at 4°C for 150 minutes at 35,000 rpm. Gradients were then immediately fractionated using the Gradient Station, and 20 x 500 µL fractions were collected in 1.5 mL Eppendorf tubes, while absorbance was monitored at 260 nm continuously. Fractions were combined in the following way: the free rRNA (F1, fractions 1 and 2), the unassembled subunits (F2, fractions 3-6), the lowly-translating monosomes (F3, fractions 7-10) and the highly-translating polysomes (F4, fractions 12-17). The pooled fractions were then concentrated using Amicon-Ultra 100K columns (Millipore), and washed two times with cold PEB. The final volume was

brought down to 200  $\mu$ L, and RNA was extracted using TRIzol reagent. Purity of the RNA was measured with a NanoDrop 2000 Spectrophotometer.

### ***In vitro* transcription of modified and unmodified RNAs**

The synthetic 'curlcake' sequences<sup>29</sup> used in this study are designed to include all possible 5-mers while minimizing the secondary RNA structure, and consist in 4 *in vitro* transcribed constructs: (i) Curlcake 1, 2244 bp; (ii) Curlcake 2, 2459 bp; (iii) Curlcake 3, 2595 bp, and (iv) Curlcake 4, 2709. The curlcake constructs were *in vitro* transcribed using Ampliscribe™ T7-Flash™ Transcription Kit (Lucigen-ASF3507) with either unmodified rNTPs (UNM), N6-methyladenosine triphosphate ( $m^6$ ATP), 5-methylcytosine triphosphate ( $m^5$ CTP), 5-hydroxymethylcytosine triphosphate ( $hm^5$ CTP) or pseudouridine triphosphate ( $\Psi$ TP). All modified NTPs were purchased from TriLink. The sequences included in the short unmodified dataset (UNM-S), which included *B. subtilis* guanine riboswitch, *B. subtilis* lysine riboswitch and *Tetrahymena* ribozyme, were also produced by *in vitro* transcription using Ampliscribe™ T7-Flash™ Transcription Kit (Lucigen-ASF3507). All constructs were 5' capped using vaccinia capping enzyme (NEB-M2080S) and polyadenylated using *E. coli* Poly(A) Polymerase (NEB-M0276S). Poly(A)-tailed RNAs were purified using RNAClean XP beads, and the addition of poly(A)-tail was confirmed using Agilent 4200 Tapestation. Concentration was determined using Qubit Fluorometric Quantitation. Purity of the IVT product was measured with NanoDrop 2000 Spectrophotometer.

### **Direct RNA library preparation and sequencing of *in vitro* transcribed constructs**

The RNA libraries for direct RNA Sequencing (SQK-RNA001) were prepared following the ONT Direct RNA Sequencing protocol version DRS\_9026\_v1\_revP\_15Dec2016, which corresponds to the flowcell FLO-MIN106. Briefly, 800 ng of Poly(A)-tailed and capped RNA (200 ng per construct) was ligated to ONT RT Adaptor (RTA) using concentrated T4 DNA Ligase (NEB-M0202T), and was reverse transcribed using SuperScript III RT (Thermo Fisher Scientific-18080044). The products were purified using 1.8X Agencourt RNAClean XP beads (Fisher Scientific-NC0068576), washing with 70% freshly prepared ethanol. RNA Adapter (RMX) was ligated onto the RNA:DNA hybrid, and the mix was purified using 1X Agencourt RNAClean XP beads, washing with Wash buffer (WSB) twice. The sample was then eluted in Elution Buffer (ELB) and mixed with RNA running buffer (RRB) prior to loading onto a primed R9.4.1 flowcell, and ran on a MinION sequencer with MinKNOW acquisition software version 1.15.1. The sequencing was performed in independent days and using a different flowcell for each sample (UNM,  $m^6$ A,  $m^5$ C,  $hm^5$ C,  $\Psi$ , UNM-S).

### **Direct RNA library preparation and sequencing of yeast total RNAs and mRNAs**

Here we performed direct RNA sequencing of two types of *S. cerevisiae* RNA inputs: i) total RNA from *S. cerevisiae*, and ii) polyA-selected RNA from *S. cerevisiae*. Yeast total RNAs were polyadenylated using *E. coli* Poly(A) Polymerase (NEB, M0276S), following the commercial protocol, prior to starting the library prep. Yeast polyA-selected RNA was directly used as input to start the libraries since they already contain poly(A) tail. Four different direct RNA libraries were barcoded according to the recent

protocol that we recently published<sup>89</sup>. Custom RT adaptors (IDT) were annealed using following conditions: custom Oligo A and B (**Table S11**) were mixed in annealing buffer (0.01 M Tris-Cl pH 7.5, 0.05M NaCl) to the final concentration of 1.4  $\mu$ M each in a total volume of 75  $\mu$ L. The mixture was incubated at 94°C for 5 minutes and slowly cooled down (-0.1°C/s) to room temperature. RNA library for direct RNA Sequencing (SQK-RNA002) was prepared following the ONT Direct RNA Sequencing protocol version DRS\_9080\_v2\_rev1\_14Aug2019 with half reaction for each library until the RNA Adapter (RMX) ligation step. Per reaction (half), 250 ng total of yeast RNAs were ligated to pre-annealed custom RT adaptors (IDT)<sup>89</sup> using concentrated T4 DNA Ligase (NEB-M0202T), and was reverse transcribed using Maxima H Minus RT (Thermo Scientific, EP0752), without the heat inactivation step. The products were purified using 1.8X Agencourt RNAClean XP beads (Fisher Scientific-NC0068576) and washed with 70% freshly prepared ethanol. 50 ng of reverse transcribed RNA from each reaction was pooled and RMX adapter, composed of sequencing adapters with motor protein, was ligated onto the RNA:DNA hybrid and the mix was purified using 1X Agencourt RNAClean XP beads, washing with Wash Buffer (WSB) twice. The sample was then eluted in Elution Buffer (EB) and mixed with RNA Running Buffer (RRB) prior to loading onto a primed R9.4.1 flowcell, and ran on a MinION sequencer with MinKNOW acquisition software version v.3.5.5.

#### **NanoCMC-seq**

CMC treatment was adapted from Schwartz et al<sup>4</sup> with minor changes. Briefly, 20  $\mu$ g total RNA was incubated in NEBNext® Magnesium RNA Fragmentation Module at 94°C for 1.5 minutes. The fragmented RNA was then incubated with either 0.3 M CMC dissolved in 100  $\mu$ L TEU buffer (50 mM Tris pH 8.5, 4 mM EDTA, 7 M Urea) or 100  $\mu$ L TEU buffer (no CMC) for 20 minutes at 37°C. Reaction was stopped with 100  $\mu$ L of Buffer A (0.3 M NaOAc and 0.1 mM EDTA, pH 5.6), 700  $\mu$ L absolute ethanol, and 1  $\mu$ L GlycoBlue (Thermo Scientific, AM9515). RNA in the stop solution was chilled on dry ice for 5 minutes, and then centrifuged at maximum speed for 15 minutes at 4°C. Supernatant was removed and the pellet was washed with 70% ethanol. After air drying for a few minutes, the pellet was dissolved in 100  $\mu$ L Buffer A and mixed with 300  $\mu$ L absolute ethanol and 1  $\mu$ L GlycoBlue. After chilling on dry ice for 5 minutes, the solution was then centrifuged at maximum speed for 15 minutes at 4°C. Supernatant was removed, and the pellet was washed with 70% ethanol. After washing, the pellet was air dried, and resuspended in 40  $\mu$ L of 50 mM sodium bicarbonate, pH 10.4, and incubated at 37°C for 3 hours. Furthermore, RNA was mixed with 100  $\mu$ L Buffer A, 700  $\mu$ L ethanol, and 1  $\mu$ L Glycoblue overnight at -20°C. The next day, the solution was centrifuged at maximum speed for 15 minutes at 4°C and the pellet was washed with 70% ethanol and dissolved in the appropriate amount of water after air drying. Unprobed and probed RNAs were treated with T4 Polynucleotide Kinase (PNK) (NEB, M0201S) as described above before proceeding with ONT Direct cDNA sequencing.

Before starting the library preparation, 9  $\mu$ L of 100  $\mu$ M Reverse-transcription primer (Original ONT VNP: 5' /5Phos/ACTTGCTGTGCTCTATCTTCTTTTTTTTTTTTTTTTTTTVN 3') and 9  $\mu$ L of 100  $\mu$ M complementary oligo (CompA: 5' GAAGATAGAGCGACAGGCAAGTA 3' ) were mixed with 1  $\mu$ L 0.2 M Tris pH 7.5 and 1  $\mu$ L 1 M NaCl. The mix was incubated at 94°C for 1 minute and the

temperature was ramped down to 25°C (-0.1°C/s) in order to pre-anneal the oligos. Then, 100 ng polyA-tailed RNA was mixed with 1 µL pre-annealed VNP+CompA, 1 µL 10 mM dNTP mix, 4 µL 5X RT Buffer, 1 µL RNasin® Ribonuclease Inhibitor (Promega, N2511), 1 µL Maxima H Minus RT (Thermo Scientific. EP0742) and nuclease-free water up to 20 µL. The reverse-transcription mix was incubated at 60°C for 60 minutes and inactivated by heating at 85°C for 5 minutes before moving onto ice. Furthermore, RNase Cocktail (Thermo Scientific, AM2286) was added to the mix in order to digest the RNA and the mix was incubated at 37°C for 10 minutes. Then the reaction was cleaned up using 1.2X AMPure XP Beads (Agencourt, A63881). In order to be able to ligate the sequencing adapters the first strand, 1 µL 100 µM CompA was again annealed to the 15 µL cDNA in a tube with 2.25 µL 0.1 M Tris pH 7.5, 2.25 µL 0.5 M NaCl and 2 µL nuclease-free water. The mix was incubated at 94°C for 1 minute and the temperature was ramped down to 25 °C (-0.1°C/s) in order to anneal the complementary to the first strand cDNA. Furthermore, 22.5 µL first strand cDNA was mixed with 2.5 µL Native Barcode (EXP-NBD104) and 25 µL Blunt/TA Ligase Mix (NEB, M0367S) and incubated in room temperature for 10 minutes. The reaction was cleaned up using 1X AMPure XP beads and the libraries were pooled into one tube that finally contains 200 fmol library. The pooled library was then ligated to the sequencing adapter (AMII) using Quick T4 DNA Ligase (NEB, M2200S) in room temperature for 10 minutes, followed with 0.65X AMPure XP Bead cleanup using ABB Buffer for washing. The sample was then eluted in Elution Buffer (EB) and mixed with Sequencing Buffer (SQB) and Loading Beads (LB) prior to loading onto a primed R9.4.1 flowcell, and ran on a MinION sequencer with MinkNOW acquisition software version v.3.5.5.

#### **Analysis of nanoCMC-seq**

Reads were base-called with stand-alone Guppy version 3.6.1 with default parameters running in GPU, with built-in demultiplexing tool of Guppy. Unclassified reads were then demultiplexed further using Porechop with --barcode\_threshold 50 option (<https://github.com/rwick/Porechop>). Then all the merged classified reads were mapped to cytosolic and mitochondrial ribosomal RNA sequences in *S. cerevisiae* using minimap2 default. Furthermore, a custom script was used to extract RT-drop signatures and the RT-drop scores were plotted using ggplot2. All scripts used to process nanoCMC-seq data with RT-Drop information have been made available in GitHub ([https://github.com/novoalab/yeast\\_RNA\\_Mod](https://github.com/novoalab/yeast_RNA_Mod)). Notably, due to the 5' end truncation of the nanopore sequencing reads by ~13 nt, RT-drop positions were shifted by 13 nt to accurately determine the exact RT-drop positions. To identify significant RT drops in a given transcript, we first computed RT-drop scores at each site, which took the difference in the coverage at a given position (0) relative to the previous position (-1). We then computed the difference (delta RT drop-off score) in RT-drop scores between CMC-probed and unprobed conditions. Lastly, we normalized the delta RT drop-off score at each position by the median RT drop-off per transcript, leading to final CMC-Scores, which can be compared across transcripts. Positions with CMC-Score greater than 25 were considered significant, i.e. to contain a pseudouridine. We should note that the nanoCMC-seq signal-to-noise ratio is dependent on the coverage of the individual transcript.

## Demultiplexing direct RNA sequencing

Demultiplexing of the barcoded direct RNA sequencing libraries was performed using DeePlexiCon with default parameters<sup>89</sup>. Reads with demultiplexing confidence scores greater than 0.95 were kept for downstream analyses. We used a lower score in the case of polysomal fractions and mRNA runs (0.8), due to the low read coverage of some fractions and/or genes. We should note that the dataset was also analyzed using 0.95 threshold, and results and conclusions of the analysis did not change, compared to those obtained using 0.80 threshold.

## Base-calling direct RNA sequencing

Reads were base-called with stand-alone Albacore versions 2.1.7 and 2.3.4 with the `--disable_filtering` parameter, and stand-alone Guppy versions 2.3.1 and 3.0.3 with default parameters running in CPU. In-house scripts were used for computing the number of unique and common base-called reads between the different approaches, as well as to compare the tendency of each base-caller regarding read lengths and qualities. Both Albacore and Guppy are available to ONT customers via their community site (<https://community.nanoporetech.com/>). Differences between the base-called features using distinct base-callers were determined using Kruskal-Wallis test with Bonferroni correction for pairwise comparisons, whereas differences between unmodified and modified sites were assessed using Mann-Whitney-Wilcoxon test.

## Mapping algorithms and parameters

Reads were mapped using either *Minimap2*<sup>45</sup> or *GraphMap*<sup>46</sup>. *Minimap2* version 2.14 was run with two different parameter settings: (i) `minimap2 -ax map-ont`, which is the recommended setting for direct RNA sequencing mapping, and thus we refer to as 'default', and (ii) `minimap2 -ax map-ont -k 5`, which we refer to as 'sensitive'. *GraphMap* version 0.5.2 was also run with two different parameter settings, for comparison, (i) `graphmap align`, using 'default' parameters, and (ii) `graphmap align --rebuild-index -v 1 --double-index --mapq -1 -x sensitive -z 1 -K fastq --min-read-len 0 -A 7 -k 5`, which is expected to increase the tolerance to errors that may occur under the presence of RNA modifications, and thus we refer to as 'sensitive'. Yeast total RNA runs were mapped to ribosomal RNAs and non-coding RNA transcripts using graphmap with default settings. Yeast poly(A)-selected runs were mapped to the yeast genome (SacCer3) using minimap2 with `-ax splice -k14 -uf` parameters. The scripts can be found in the GitHub repository [https://github.com/novoalab/yeast\\_RNA\\_Mod](https://github.com/novoalab/yeast_RNA_Mod). Sequencing, base-calling and mapping statistics for all yeast sequencing runs (total RNA and polyA-selected RNA) can be found in **Tables S12** and **13**.

## Analysis of base-called features in curlcakes

Sam files were transformed into bam files using Samtools version 1.9<sup>90</sup>, and were then sorted and indexed in order to visualize the data using the Integrative Genomics Viewer (IGV) version 2.4.16<sup>91</sup>. Base-called features were extracted with *EpiNano* version 1.1 (<https://github.com/enovoa/EpiNano>). Principal Component Analysis (PCA) was used to reduce the dimensionality of the base-calling error data to visually inspect for base-calling differences, using as input the base-called features (mismatch

frequency, deletion frequency and per-base quality) from all 5 positions of each k-mer. Only k-mers that contained a given modification once in the 5-mer were included in the analysis. All scripts used to analyze *in vitro* transcribed sequences using different base-calling algorithms and mappers, as well as to generate the Figures related to their analysis are available in [https://github.com/novoalab/Best\\_Practices\\_dRNAseq\\_analysis](https://github.com/novoalab/Best_Practices_dRNAseq_analysis).

### **Analysis of base-called features in yeast RNAs**

Sam files were transformed into bam files using Samtools version 1.9<sup>90</sup>, then sorted and indexed in order to visualize the data using the Integrative Genomics Viewer (IGV) version 2.4.16<sup>91</sup>. Base-called features were extracted using *EpiNano* version 1.1 with minor modifications, which consisted in including in the output csv file the directionality of mismatched bases (C\_frequency, G\_frequency, A\_frequency, U\_frequency). The modified *EpiNano* script can be found at [https://github.com/novoalab/yeast\\_RNA\\_Mod](https://github.com/novoalab/yeast_RNA_Mod). Scripts for the analysis and visualization of base-called features are also included in the same GitHub repository.

### **Visualization per-read current intensities using Nanopolish**

*Nanopolish* eventalign output was processed to extract the current intensity values corresponding to the 15-mer regions centered in the modified sites, for the following sites: (i) 6  $\Psi$  rRNA sites for which knockout data was available (25s:2133, 25s:2129, 25s:2826, 25s:2880, 25s:2264, 18s:1187), for all 4 sequencing datasets (wild type, snR3-KO, snR34-KO, snR36-KO); (ii) 4 Nm sites for which knockout data was available (25s:817, 25s:908, 25s:1133, 25s:1888), for all 4 sequencing datasets (wild type, snR60-KO, snR61-KO, snR62-KO); (iii) 7  $\Psi$  snRNA/snoRNA sites which were identified as heat-sensitive, for which there was a minimum of 100 reads of coverage. Reads with empty values in the 15-mer region in the *Nanopolish* eventalign output were omitted from the analysis.

### **Analysis of current intensity, dwell time and trace**

In this work, we used two different softwares to extract current intensity: Nanopolish<sup>92</sup> and Tombo<sup>50</sup>. Nanopolish was used to extract the aligned current intensity values per read and position, using the option `--scale-events`. Mean current intensity per-position was computed by summing the current intensities of all reads aligned to the same position, divided by the total number of reads mapping at a given position. All scripts used to process *Nanopolish* event align output, including scripts to display mean current intensity values along transcripts have been made available in GitHub (<https://github.com/novoalab/nanoRMS>).

Signal intensity, dwell time and trace were retrieved using `get_features.py` script, which is available as part of *nanoRMS*. This program internally uses: minimap2 (read alignment), Tombo (calculation of signal intensity and dwell time) and ont-fast5-api (retrieval of trace). Trace represents the probability that a given signal intensity chunk may be originating from each of the 4 canonical bases (A, C, G and T/U), and it is reported relative to the reference base. For example, in a T reference position that is incorrectly reported as C (common base-calling error observed for  $\Psi$  sites), the trace value will be

reported for the reference base (T in this case). Then, the final read alignment and all the features are stored into sorted BAM files. All scripts necessary to retrieve and store per-read, per-position features and plot/calculate results are available within the *nanoRMS* GitHub repository (<https://github.com/novoalab/nanoRMS>).

### ***De novo* prediction of pseudouridine modifications on yeast mitochondrial rRNAs**

To systematically identify  $\Psi$  sites *de novo* based on the  $\Psi$  base-calling signatures, we first extracted the mismatch frequency and per-base mismatch frequency (C\_freq, A\_freq, U\_freq, G\_freq) from both unmodified (U) and modified ( $\Psi$ ) sites from cytosolic ribosomal RNAs, from three biological replicates. As expected, C mismatch frequency (C\_freq) and global mismatch frequency (mis\_freq) showed clearly distinct distributions when comparing unmodified and  $\Psi$ -modified sites (**Figure 5A**). We then determined the optimal cut-points for these two features using the *cutpointR* package in R with *oc\_youden\_kernel* method, which applies Kernel smoothing and maximizes the Youden-Indexing. This approach predicted C\_freq=0.137 and mis\_freq= 0.587 as optimal cut-offs. For the mitochondrial ribosomal RNA, we filtered the uridine sites based on the selected features and assigned those that are replicable in three biological replicates as “candidate” pseudouridine sites.

### ***De novo* prediction of pseudouridine modifications in yeast mRNAs and non-coding RNAs**

Due to the lower stoichiometry of modification of noncoding RNAs (snRNA and snoRNAs) and mRNAs, we focused on analysis of the *de novo* detection of  $\Psi$  sites whose pseudouridylation levels would be changing between two conditions, either by comparing normal and stress (heat-shock) conditions, or by comparing the base-calling ‘error’ patterns of wild type strains and Pus1 or Pus4-deficient strains. Only sites which passed the coverage filter ( $n > 30$  reads) in both biological replicates from both conditions were considered in the analysis (**Table S6**). Sites with minimal mismatch frequency difference of 0.1 between the two conditions in both replicates that met the identified  $\Psi$  signature (C\_freq=0.137 and mis\_freq= 0.587) were considered as true  $\Psi$  sites that were either heat-sensitive, Pus1-dependent, or Pus4-dependent, respectively. The individual candidate  $\Psi$  mRNA and ncRNA sites identified using nanopore sequencing, as well as the previously reported  $\Psi$  mRNA and ncRNA sites (using CMC probing coupled to Illumina sequencing) can be found in **Tables S7-S10**.

### **Prediction of RNA modification stoichiometry using *nanoRMS***

Per-position features from individual reads were stored in BAM files using *pysam* (<https://github.com/pysam-developers/pysam>) and stored them either in Numpy arrays (<https://numpy.org/>) or Pandas DataFrames (<https://pandas.pydata.org/>) using the script *get\_features.py*, which is available as part of *nanoRMS*. Models were trained with combinations of features with diverse ranges of sequence contexts surrounding the modified sites ( $k=1-15$ ). Features used to predict stoichiometry included: (i) current intensity (SI), (ii) dwell time in the centre of the pore (at position 0, DT/DT0), (iii) dwell time at helicase centre (shifted by 10 positions, DT10) and (iv) base probability (trace, TR). Estimation of modification frequency was performed using unsupervised (GMM, KMEANS, IsolationForest, OneClassSVM) and supervised (KNN, RandomForest) machine

learning methods implemented in sklearn (<https://sklearn.org/>). Plots were built using matplotlib and seaborn (<https://seaborn.pydata.org/>).

Trained models were first benchmarked with unmodified (KO) and modified (WT) reads from rRNA mutants dataset, to identify which machine learning methods and which combination of features discriminated between modified and unmodified reads. Then, we tested how the diverse models would perform at diverse stoichiometries of modification. To this end, we simulated samples with varying levels of modification: 0%, 20%, 40%, 60%, 80% and 100% (using mixes of KO and WT reads) and estimated the modification level in those simulated samples by comparing them to KO (Figure S5C).

*NanoRMS* performed best when trained with signal intensity (SI) + trace (TR) as features, and when using KNN supervised models or KMEANS unsupervised models, both for  $\Psi$  and Nm-modified sites. Predictions by each clustering algorithm, and for each individual rRNA modified site, are shown in Table S4. For mRNA and ncRNA analysis, only sites with more than 30 reads of coverage in all conditions and replicates were included for predicting RNA modification stoichiometry. Prediction of RNA modification stoichiometry in mRNAs and non-coding RNAs was performed using signal intensity + trace as features, and k-means as classification algorithm. Stoichiometry changes were reported as the difference in predicted stoichiometry between the two conditions. All code and examples to predict RNA modification stoichiometry are available as part of the *nanoRMS* GitHub repository (<https://github.com/novoalab/nanoRMS>).

#### DATA AVAILABILITY

For *in vitro* transcribed datasets, FAST5 files used in this work were already publicly available (UNM and m<sup>6</sup>A: PRJNA511582), or have been made publicly available in SRA (m<sup>5</sup>C: PRJNA563591; hm<sup>5</sup>C: PRJNA548268;  $\Psi$ : PRJNA549001, UNM-S: PRJNA575545,  $\Psi$  with varying stoichiometries: PRJNA695584). Base-called and demultiplexed FASTQ from all yeast RNA direct RNA sequencing data runs have been made publicly available in GEO, under the accession number GSE148603, including processed *EpiNano* outputs. FAST5 files for all yeast RNA direct RNA sequencing are available in ENA under accession PRJEB37798 and PRJEB41495. A detailed description of the datasets used and sequenced in this work, with their corresponding GEO and ENA/SRA IDs can be found in Table S14.

#### CODE AVAILABILITY

All scripts and code used in this work have been made available in GitHub: [https://github.com/novoalab/Best\\_Practices\\_dRNAseq\\_analysis](https://github.com/novoalab/Best_Practices_dRNAseq_analysis) (analysis of *in vitro* curlicake datasets), [https://github.com/novoalab/yeast\\_RNA\\_Mod](https://github.com/novoalab/yeast_RNA_Mod) (analysis of *in vivo* datasets) and <https://github.com/novoalab/nanoRMS> (prediction of RNA modifications and estimation of RNA modification stoichiometries).

## ACKNOWLEDGEMENTS

We thank all the members of the Novoa lab for their valuable insights and discussion. We thank Vivek Malhotra for sharing the Pus1 and Pus4 knockout strains. OB is supported by a UNSW International PhD fellowship. MCL is supported by an FPI Severo-Ochoa fellowship by the Spanish Ministry of Economy, Industry and Competitiveness (MEIC). IM and SC are supported by "la Caixa" INPhINIT PhD fellowships (LCF/BQ/DI18/11660028 and LCF/BQ/DI19/11730036, respectively). This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 713673. This work was supported by the Australian Research Council (DP180103571 to EMN) and the Spanish Ministry of Economy, Industry and Competitiveness (MEIC) (PGC2018-098152-A-100 to EMN). We acknowledge the support of the MEIC to the EMBL partnership, Centro de Excelencia Severo Ochoa and CERCA Programme/Generalitat de Catalunya.

## AUTHOR CONTRIBUTIONS

OB and MCL performed the majority of wet lab experiments, including RNA extraction and nanopore library preparation. OB and LPP performed bioinformatic analysis of the data, together with JMR and EMN. OB conceived and performed nanoCMC-Seq experiments. MCL produced the *in vitro* transcribed sequences with modifications and their corresponding nanopore libraries. OB prepared and sequenced the *in vitro* transcribed sequences with different pseudouridine stoichiometries. LPP benchmarked and wrote the *nanoRMS* code, together with OB and EMN. JMR performed bioinformatic analyses on *in vitro* transcribed constructs and compared base-calling and mapping algorithms. IM built polysome gradients and helped with their corresponding nanopore libraries. SC and IM prepared and sequenced the 2'-O-methylation mutant strains. RM and HGVS cultured the *S. cerevisiae* strains under different stress conditions. HL contributed with code for the analysis of current intensity values. ASC and SS cultured all snoRNA-depleted yeast mutant strains and extracted their total RNA. EMN conceived the project. EMN supervised the work, with the assistance of JSM. MCL, OB and EMN built the figures. OB, MCL and EMN wrote the paper, with contributions from all authors.

## DECLARATIONS OF INTERESTS

EMN has received travel and accommodation expenses to speak at Oxford Nanopore Technologies conferences. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## REFERENCES

1. Dominissini, D. *et al.* Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature* **485**, 201–206 (2012).
2. Meyer, K. D. *et al.* Comprehensive Analysis of mRNA Methylation Reveals Enrichment in 3'

- 1072 UTRs and near Stop Codons. *Cell* vol. 149 1635–1646 (2012).
- 1073 3. Carlile, T. M. *et al.* Pseudouridine profiling reveals regulated mRNA pseudouridylation in yeast  
1074 and human cells. *Nature* **515**, 143–146 (2014).
- 1075 4. Schwartz, S. *et al.* Transcriptome-wide mapping reveals widespread dynamic-regulated  
1076 pseudouridylation of ncRNA and mRNA. *Cell* **159**, 148–162 (2014).
- 1077 5. Lovejoy, A. F., Riordan, D. P. & Brown, P. O. Transcriptome-wide mapping of pseudouridines:  
1078 pseudouridine synthases modify specific mRNAs in *S. cerevisiae*. *PLoS One* **9**, e110799 (2014).
- 1079 6. Li, X. *et al.* Chemical pulldown reveals dynamic pseudouridylation of the mammalian  
1080 transcriptome. *Nat. Chem. Biol.* **11**, 592–597 (2015).
- 1081 7. Hussain, S., Aleksic, J., Blanco, S., Dietmann, S. & Frye, M. Characterizing 5-methylcytosine in  
1082 the mammalian epitranscriptome. *Genome Biol.* **14**, 215 (2013).
- 1083 8. Huang, T., Chen, W., Liu, J., Gu, N. & Zhang, R. Genome-wide identification of mRNA 5-  
1084 methylcytosine in mammals. *Nat. Struct. Mol. Biol.* **26**, 380–388 (2019).
- 1085 9. Delatte, B. *et al.* RNA biochemistry. Transcriptome-wide distribution and function of RNA  
1086 hydroxymethylcytosine. *Science* **351**, 282–285 (2016).
- 1087 10. Safra, M. *et al.* The m1A landscape on cytosolic and mitochondrial mRNA at single-base  
1088 resolution. *Nature* **551**, 251–255 (2017).
- 1089 11. Li, X. *et al.* Base-Resolution Mapping Reveals Distinct mA Methylome in Nuclear- and  
1090 Mitochondrial-Encoded Transcripts. *Mol. Cell* **68**, 993–1005.e9 (2017).
- 1091 12. Marchand, V. *et al.* AlkAniline-Seq: Profiling of m7G and m3C RNA Modifications at Single  
1092 Nucleotide Resolution. *Angew. Chem. Int. Ed.* **57**, 16785–16790 (2018).
- 1093 13. Arango, D. *et al.* Acetylation of Cytidine in mRNA Promotes Translation Efficiency. *Cell* **175**,  
1094 1872–1886.e24 (2018).
- 1095 14. Sas-Chen, A. *et al.* Dynamic RNA acetylation revealed by quantitative cross-evolutionary  
1096 mapping. *Nature* (2020) doi:10.1038/s41586-020-2418-2.
- 1097 15. Zhang, L.-S. *et al.* Transcriptome-wide Mapping of Internal N7-Methylguanosine Methylome in  
1098 Mammalian mRNA. *Mol. Cell* **74**, 1304–1316.e8 (2019).
- 1099 16. Pandolfini, L. *et al.* METTL1 Promotes let-7 MicroRNA Processing via m7G Methylation. *Mol. Cell*  
1100 **74**, 1278–1290.e9 (2019).
- 1101 17. Delaunay, S. & Frye, M. RNA modifications regulating cell fate in cancer. *Nat. Cell Biol.* **21**, 552–

1102 559 (2019).

1103 18. Haussmann, I. U. *et al.* m6A potentiates Sxl alternative pre-mRNA splicing for robust *Drosophila*  
 1104 sex determination. *Nature* vol. 540 301–304 (2016).

1105 19. Vu, L. P. *et al.* The N6-methyladenosine (m6A)-forming enzyme METTL3 controls myeloid  
 1106 differentiation of normal hematopoietic and leukemia cells. *Nature Medicine* vol. 23 1369–1376  
 1107 (2017).

1108 20. Novoa, E. M., Mason, C. E. & Mattick, J. S. Charting the unknown epitranscriptome. *Nat. Rev.*  
 1109 *Mol. Cell Biol.* **18**, 339–340 (2017).

1110 21. Anreiter, I., Mir, Q., Simpson, J. T., Janga, S. C. & Soller, M. New Twists in Detecting mRNA  
 1111 Modification Dynamics. *Trends Biotechnol.* **0**, (2020).

1112 22. Li, X., Xiong, X. & Yi, C. Epitranscriptome sequencing technologies: decoding RNA modifications.  
 1113 *Nat. Methods* **14**, 23–31 (2016).

1114 23. Motorin, Y. & Helm, M. Methods for RNA Modification Mapping Using Deep Sequencing:  
 1115 Established and New Emerging Technologies. *Genes* **10**, (2019).

1116 24. Grozhik, A. V. *et al.* Antibody cross-reactivity accounts for widespread appearance of m<sup>1</sup>A in  
 1117 5'UTRs. *Nat. Commun.* **10**, 1–13 (2019).

1118 25. Lahens, N. F. *et al.* IVT-seq reveals extreme bias in RNA sequencing. *Genome Biol.* **15**, R86  
 1119 (2014).

1120 26. Garalde, D. R. *et al.* Highly parallel direct RNA sequencing on an array of nanopores. *Nat.*  
 1121 *Methods* **15**, 201–206 (2018).

1122 27. Jonkhout, N. *et al.* The RNA modification landscape in human disease. *RNA* **23**, 1754–1769  
 1123 (2017).

1124 28. Leger, A., Amaral, P. P., Pandolfini, L. & Capitanchik, C. RNA modifications detection by  
 1125 comparative Nanopore direct RNA sequencing. *BioRxiv* (2019).

1126 29. Liu, H. *et al.* Accurate detection of m6A RNA modifications in native RNA sequences. *Nat.*  
 1127 *Commun.* **10**, 4079 (2019).

1128 30. Parker, M. T. *et al.* Nanopore direct RNA sequencing maps the complexity of Arabidopsis mRNA  
 1129 processing and m6A modification. *eLife* vol. 9 (2020).

1130 31. Price, A. M. *et al.* Direct RNA sequencing reveals m6A modifications on adenovirus RNA are  
 1131 necessary for efficient splicing. *bioRxiv* 865485 (2019) doi:10.1101/865485.

- 1132 32. Wongsurawat, T. *et al.* Decoding the Epitranscriptional Landscape from Native RNA Sequences.  
1133 *bioRxiv* 487819 (2018) doi:10.1101/487819.
- 1134 33. Lorenz, D. A., Sathe, S., Einstein, J. M. & Yeo, G. W. Direct RNA sequencing enables m6A  
1135 detection in endogenous transcript isoforms at base-specific resolution. *RNA* **26**, 19–28 (2020).
- 1136 34. Pratanwanich, P. N. *et al.* Detection of differential RNA modifications from direct RNA sequencing  
1137 of human cell lines. *bioRxiv* 2020.06.18.160010 (2020) doi:10.1101/2020.06.18.160010.
- 1138 35. Jack, K. *et al.* rRNA pseudouridylation defects affect ribosomal ligand binding and translational  
1139 fidelity from yeast to human cells. *Mol. Cell* **44**, 660–666 (2011).
- 1140 36. Yoon, A. *et al.* Impaired control of IRES-mediated translation in X-linked dyskeratosis congenita.  
1141 *Science* **312**, 902–906 (2006).
- 1142 37. Bellodi, C. *et al.* Loss of Function of the Tumor Suppressor DKC1 Perturbs p27 Translation  
1143 Control and Contributes to Pituitary Tumorigenesis. *Cancer Research* vol. 70 6026–6035 (2010).
- 1144 38. Wu, G., Xiao, M., Yang, C. & Yu, Y.-T. U2 snRNA is inducibly pseudouridylated at novel sites by  
1145 Pus7p and snR81 RNP. *EMBO J.* **30**, 79–89 (2011).
- 1146 39. Taoka, M., Nobe, Y., Hori, M. & Takeuchi, A. A mass spectrometry-based method for  
1147 comprehensive quantitative determination of post-transcriptional RNA modifications: the complete  
1148 chemical structure of .... *Nucleic acids* (2015).
- 1149 40. Basu, A. *et al.* Requirement of rRNA methylation for 80S ribosome assembly on a cohort of  
1150 cellular internal ribosome entry sites. *Mol. Cell. Biol.* **31**, 4482–4499 (2011).
- 1151 41. Marcel, V. *et al.* p53 acts as a safeguard of translational control by regulating fibrillarin and rRNA  
1152 methylation in cancer. *Cancer Cell* **24**, 318–330 (2013).
- 1153 42. Belin, S. *et al.* Dysregulation of Ribosome Biogenesis and Translational Capacity Is Associated  
1154 with Tumor Progression of Human Breast Cancer Cells. *PLoS ONE* vol. 4 e7147 (2009).
- 1155 43. Buchhaupt, M. *et al.* Partial methylation at Am100 in 18S rRNA of baker's yeast reveals ribosome  
1156 heterogeneity on the level of eukaryotic rRNA modification. *PLoS One* **9**, e89640 (2014).
- 1157 44. Chen, H. *et al.* METTL5, an 18S rRNA-specific m6A methyltransferase, modulates expression of  
1158 stress response genes. *bioRxiv* 2020.04.27.064162 (2020) doi:10.1101/2020.04.27.064162.
- 1159 45. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* vol. 34 3094–3100  
1160 (2018).
- 1161 46. Sović, I. *et al.* Fast and sensitive mapping of nanopore sequencing reads with GraphMap. *Nat.*

1162        *Commun.* **7**, 11307 (2016).

1163    47. Liu, Q. *et al.* Detection of DNA base modifications by deep recurrent neural network on Oxford  
 1164        Nanopore sequencing data. *Nat. Commun.* **10**, 2449 (2019).

1165    48. McIntyre, A. B. R. *et al.* Single-molecule sequencing detection of N6-methyladenine in microbial  
 1166        reference materials. *Nat Commun* **10**: 579. (2019).

1167    49. De Coster, W., Stovner, E. B. & Strazisar, M. Methplotlib: analysis of modified nucleotides from  
 1168        nanopore sequencing. *Bioinformatics* **36**, 3236–3238 (2020).

1169    50. Stoiber, M. *et al.* De novo Identification of DNA Modifications Enabled by Genome-Guided  
 1170        Nanopore Signal Processing. *bioRxiv* 094672 (2017) doi:10.1101/094672.

1171    51. Taoka, M. *et al.* The complete chemical structure of *Saccharomyces cerevisiae* rRNA: partial  
 1172        pseudouridylation of U2345 in 25S rRNA by snoRNA snR9. *Nucleic Acids Res.* **44**, 8951–8961  
 1173        (2016).

1174    52. Pintard, L., Bujnicki, J. M., Lapeyre, B. & Bonnerot, C. MRM2 encodes a novel yeast  
 1175        mitochondrial 21S rRNA methyltransferase. *EMBO J.* **21**, 1139–1147 (2002).

1176    53. Sharma, S. & Lafontaine, D. L. J. 'View From A Bridge': A New Perspective on Eukaryotic rRNA  
 1177        Base Modification. *Trends Biochem. Sci.* **40**, 560–575 (2015).

1178    54. Taoka, M. *et al.* Landscape of the complete RNA chemical modifications in the human 80S  
 1179        ribosome. *Nucleic Acids Res.* **46**, 9289–9298 (2018).

1180    55. Fischer, N. *et al.* Structure of the *E. coli* ribosome--EF-Tu complex at < 3 Å resolution by C s-  
 1181        corrected cryo-EM. *Nature* **520**, 567–570 (2015).

1182    56. Sergeeva, O. V., Bogdanov, A. A. & Sergiev, P. V. What do we know about ribosomal RNA  
 1183        methylation in *Escherichia coli*? *Biochimie* **117**, 110–118 (2015).

1184    57. Sloan, K. E. *et al.* Tuning the ribosome: The influence of rRNA modification on eukaryotic  
 1185        ribosome biogenesis and function. *RNA Biol.* **14**, 1138–1152 (2017).

1186    58. Hebras, J., Krogh, N., Marty, V., Nielsen, H. & Cavaillé, J. Developmental changes of rRNA  
 1187        ribose methylations in the mouse. *RNA Biol.* 1–15 (2019).

1188    59. Higa-Nakamine, S. *et al.* Loss of ribosomal RNA modification causes developmental defects in  
 1189        zebrafish. *Nucleic Acids Res.* **40**, 391–398 (2012).

1190    60. Sahoo, T. *et al.* Prader-Willi phenotype caused by paternal deficiency for the HBII-85 C/D box  
 1191        small nucleolar RNA cluster. *Nat. Genet.* **40**, 719–721 (2008).

- 1192 61. Heiss, N. S. *et al.* X-linked dyskeratosis congenita is caused by mutations in a highly conserved  
1193 gene with putative nucleolar functions. *Nat. Genet.* **19**, 32–38 (1998).
- 1194 62. Knight, S. W. *et al.* X-Linked Dyskeratosis Congenita Is Predominantly Caused by Missense  
1195 Mutations in the DKC1 Gene. *The American Journal of Human Genetics* vol. 65 50–58 (1999).
- 1196 63. Liao, J. *et al.* Small nucleolar RNA signatures as biomarkers for non-small-cell lung cancer. *Mol.*  
1197 *Cancer* **9**, 198 (2010).
- 1198 64. Mei, Y.-P. *et al.* Small nucleolar RNA 42 acts as an oncogene in lung tumorigenesis. *Oncogene*  
1199 **31**, 2794–2804 (2012).
- 1200 65. Bortolin-Cavaille, M.-L., -L. Bortolin-Cavaille, M. & Cavaille, J. The SNORD115 (H/MBII-52) and  
1201 SNORD116 (H/MBII-85) gene clusters at the imprinted Prader-Willi locus generate canonical box  
1202 C/D snoRNAs. *Nucleic Acids Research* vol. 40 6800–6807 (2012).
- 1203 66. Eralles, J. *et al.* Evidence for rRNA 2'-O-methylation plasticity: Control of intrinsic translational  
1204 capabilities of human ribosomes. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 12934–12939 (2017).
- 1205 67. Krogh, N. *et al.* Profiling of 2'-O-Me in human rRNA reveals a subset of fractionally modified  
1206 positions and provides evidence for ribosome heterogeneity. *Nucleic Acids Research* vol. 44  
1207 7884–7895 (2016).
- 1208 68. Birkedal, U. *et al.* Profiling of ribose methylations in RNA by high-throughput sequencing. *Angew.*  
1209 *Chem. Int. Ed Engl.* **54**, 451–455 (2015).
- 1210 69. Natchiar, S. K., Myasnikov, A. G., Kratzat, H., Hazemann, I. & Klaholz, B. P. Visualization of  
1211 chemical modifications in the human 80S ribosome structure. *Nature* **551**, 472–477 (2017).
- 1212 70. van der Feltz, C., DeHaven, A. C. & Hoskins, A. A. Stress-induced Pseudouridylation Alters the  
1213 Structural Equilibrium of Yeast U2 snRNA Stem II. *J. Mol. Biol.* **430**, 524–536 (2018).
- 1214 71. Frye, M., Harada, B. T., Behm, M. & He, C. RNA modifications modulate gene expression during  
1215 development. *Science* **361**, 1346–1349 (2018).
- 1216 72. Roundtree, I. A., Evans, M. E., Pan, T. & He, C. Dynamic RNA Modifications in Gene Expression  
1217 Regulation. *Cell* **169**, 1187–1200 (2017).
- 1218 73. Li, S. & Mason, C. E. The pivotal regulatory landscape of RNA modifications. *Annu. Rev.*  
1219 *Genomics Hum. Genet.* **15**, 127–150 (2014).
- 1220 74. Wang, X. *et al.* LARP7-Mediated U6 snRNA Modification Ensures Splicing Fidelity and  
1221 Spermatogenesis in Mice. *Mol. Cell* **77**, 999–1013.e6 (2020).

1222 75. Louloup, A., Ntini, E., Conrad, T. & Ørom, U. A. V. Transient N-6-Methyladenosine  
1223 Transcriptome Sequencing Reveals a Regulatory Role of m6A in Splicing Efficiency. *Cell Reports*  
1224 vol. 23 3429–3437 (2018).

1225 76. Zhou, K. I. *et al.* Regulation of Co-transcriptional Pre-mRNA Splicing by m6A through the Low-  
1226 Complexity Protein hnRNPG. *Mol. Cell* **76**, 70–81.e9 (2019).

1227 77. Lee, Y., Choe, J., Park, O. H. & Kim, Y. K. Molecular Mechanisms Driving mRNA Degradation by  
1228 m6A Modification. *Trends Genet.* **36**, 177–188 (2020).

1229 78. Guo, M., Liu, X., Zheng, X., Huang, Y. & Chen, X. m6A RNA Modification Determines Cell Fate  
1230 by Regulating mRNA Degradation. *Cellular Reprogramming* vol. 19 225–231 (2017).

1231 79. Geula, S. *et al.* Stem cells. m6A mRNA methylation facilitates resolution of naïve pluripotency  
1232 toward differentiation. *Science* **347**, 1002–1006 (2015).

1233 80. Weng, H. *et al.* METTL14 Inhibits Hematopoietic Stem/Progenitor Differentiation and Promotes  
1234 Leukemogenesis via mRNA m6A Modification. *Cell Stem Cell* **22**, 191–205.e9 (2018).

1235 81. Lence, T. *et al.* m6A modulates neuronal functions and sex determination in Drosophila. *Nature*  
1236 **540**, 242–247 (2016).

1237 82. Kan, L. *et al.* The m6A pathway facilitates sex determination in Drosophila. *Nat. Commun.* **8**,  
1238 15737 (2017).

1239 83. Schaefer, M., Kapoor, U. & Jantsch, M. F. Understanding RNA modifications: the promises and  
1240 technological bottlenecks of the ‘epitranscriptome’. *Open Biology* vol. 7 170077 (2017).

1241 84. Depledge, D. P. *et al.* Direct RNA sequencing on nanopore arrays redefines the transcriptional  
1242 complexity of a viral pathogen. *Nat. Commun.* **10**, 754 (2019).

1243 85. Roach, N. P. *et al.* The full-length transcriptome of *C. elegans* using direct RNA sequencing.  
1244 *Genome Res.* **30**, 299–312 (2020).

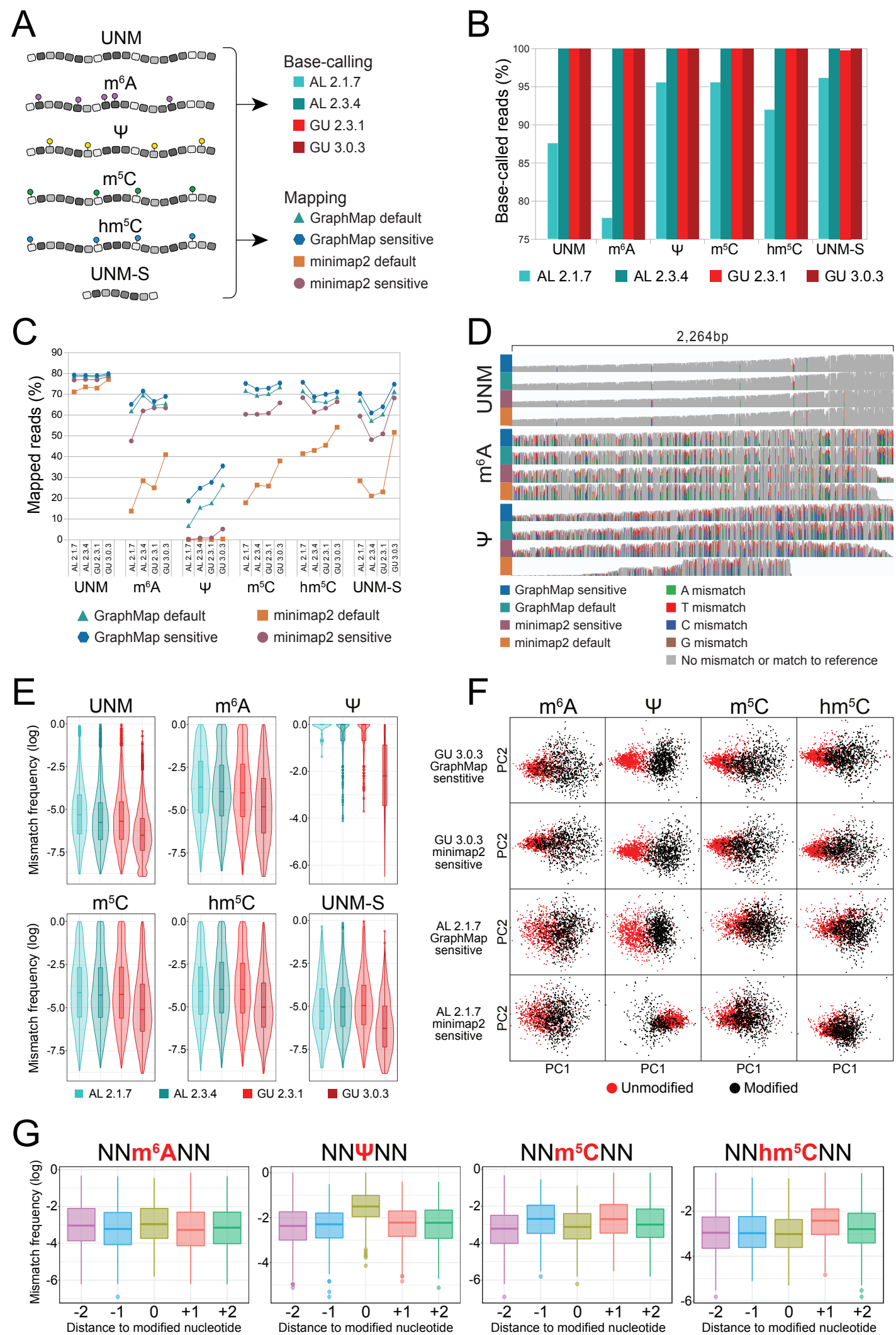
1245 86. Workman, R. E. *et al.* Nanopore native RNA sequencing of a human poly(A) transcriptome.  
1246 doi:10.1101/459529.

1247 87. Kim, D. *et al.* The Architecture of SARS-CoV-2 Transcriptome. *Cell* **181**, 914–921.e10 (2020).

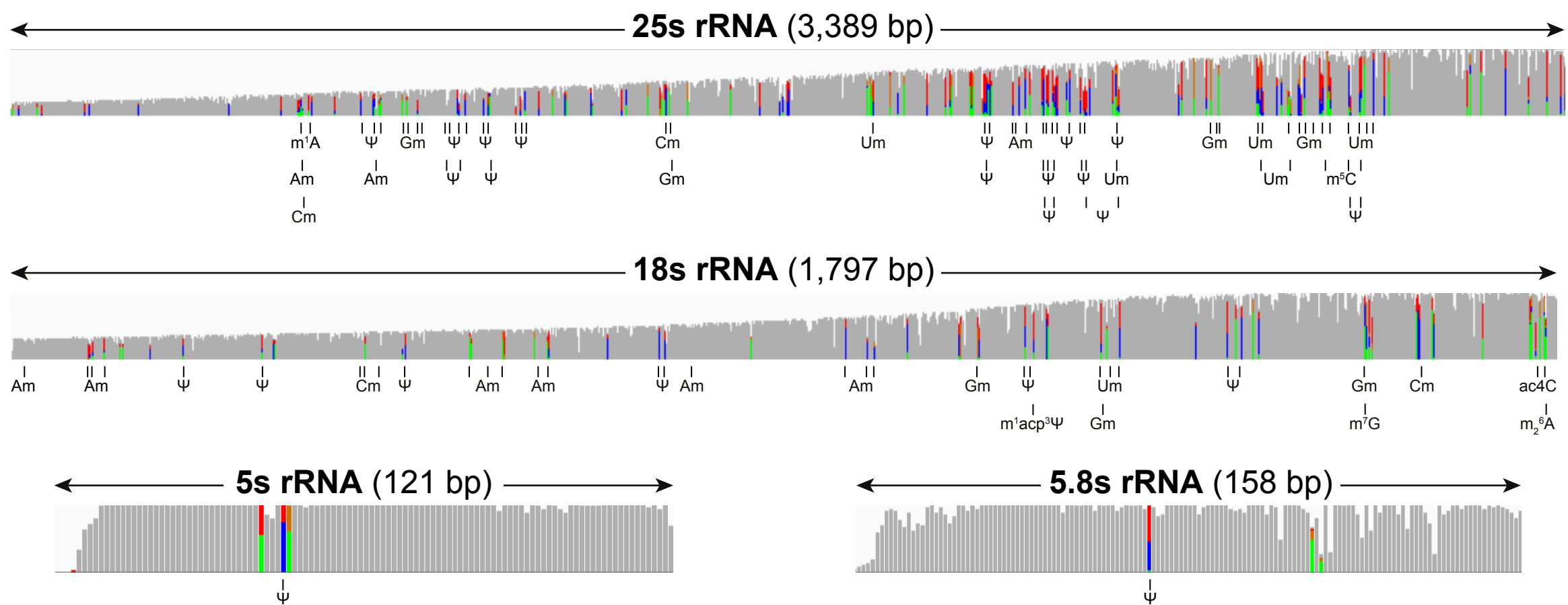
1248 88. Parker, S. *et al.* A resource for functional profiling of noncoding RNA in the yeast *Saccharomyces*  
1249 *cerevisiae*. *RNA* **23**, 1166–1171 (2017).

1250 89. Smith, M. A. *et al.* Barcoding and demultiplexing Oxford Nanopore native RNA sequencing reads  
1251 with deep residual learning. *bioRxiv* 864322 (2019) doi:10.1101/864322.

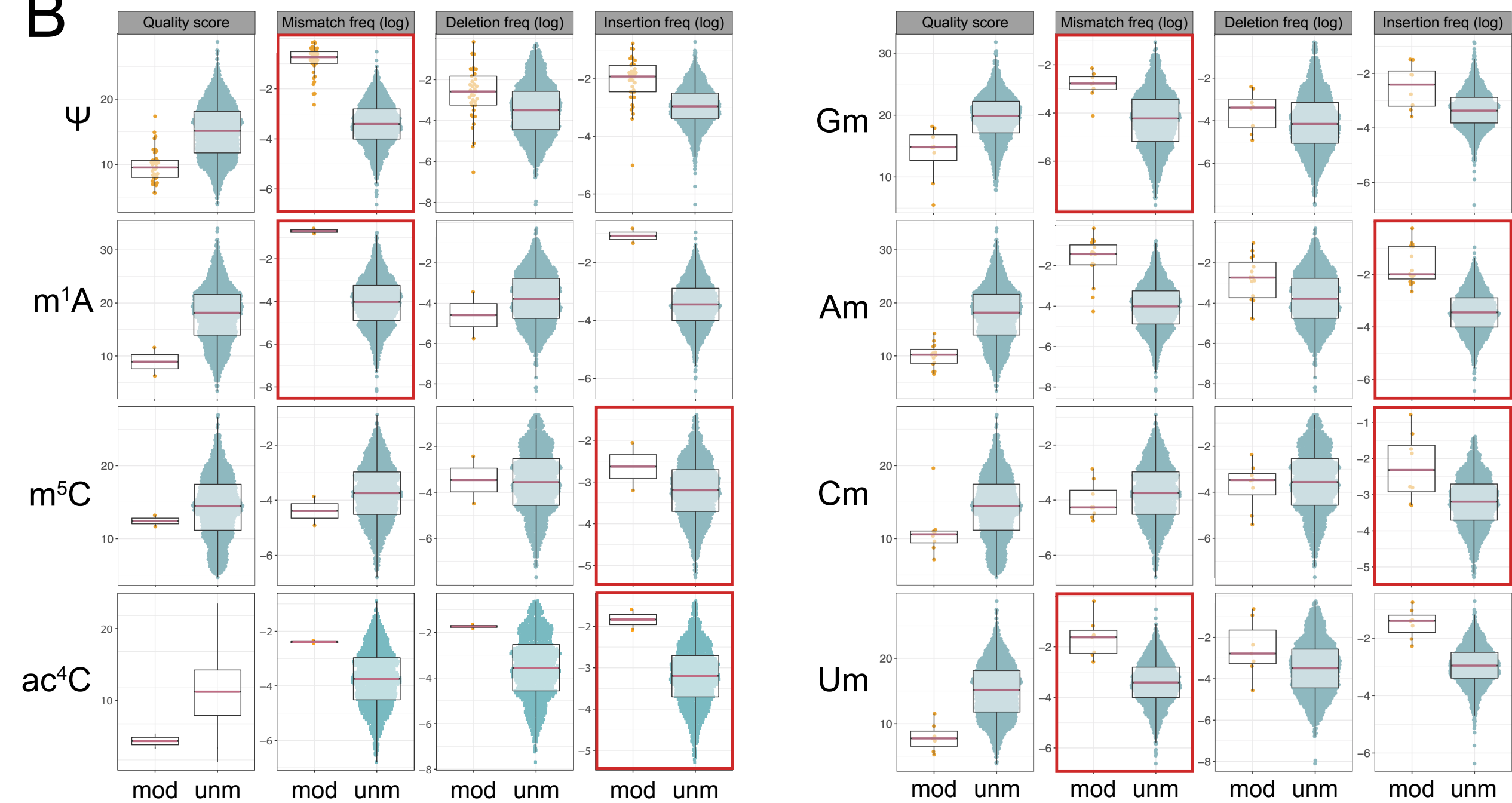
- 1252 90. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* vol. 25 2078–  
1253 2079 (2009).
- 1254 91. Robinson, J. T. *et al.* Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).
- 1255 92. Loman, N. J., Quick, J. & Simpson, J. T. A complete bacterial genome assembled de novo using  
1256 only nanopore sequencing data. *Nat. Methods* **12**, 733–735 (2015).
- 1257



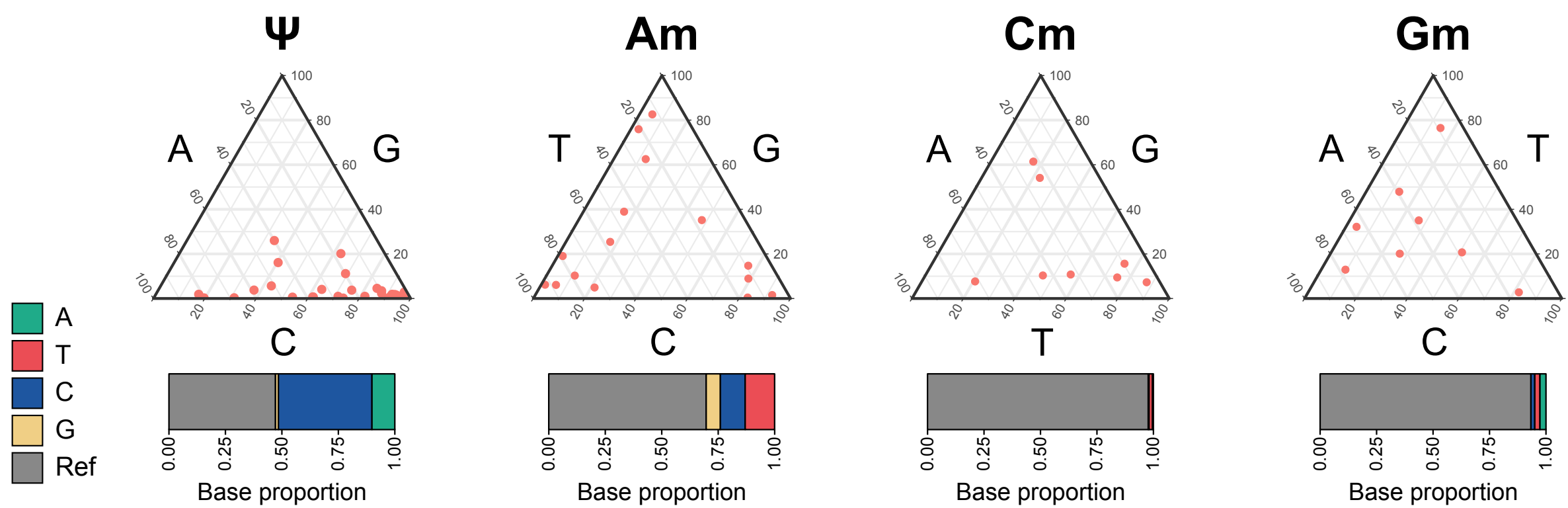
A

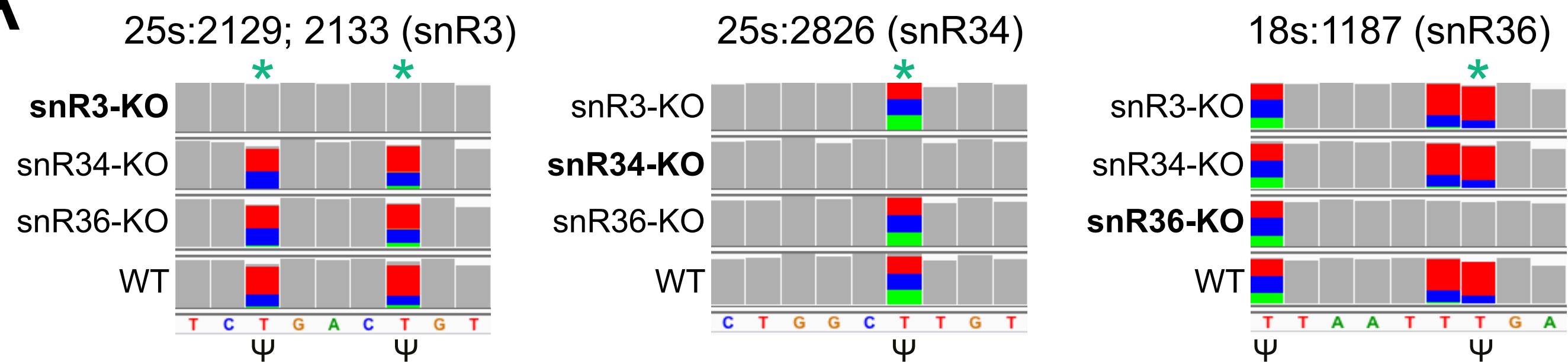
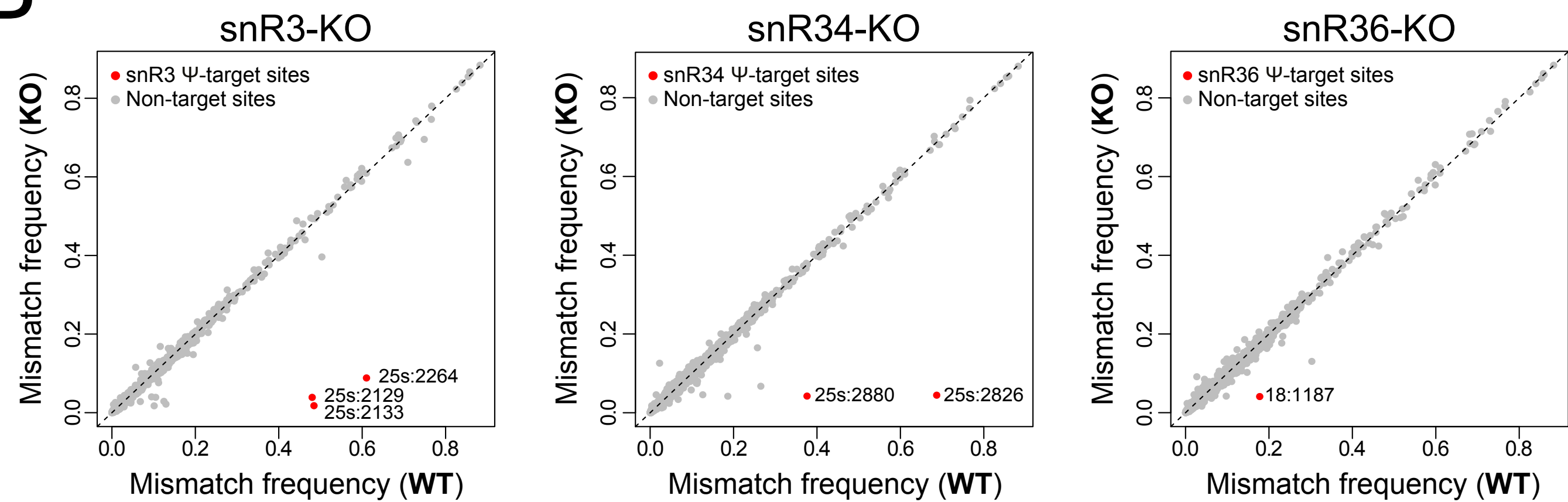
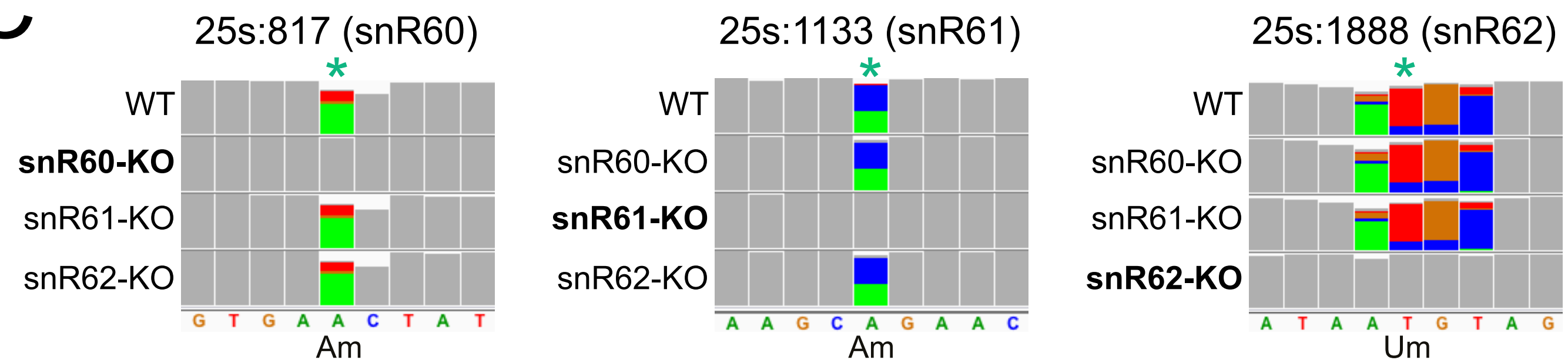
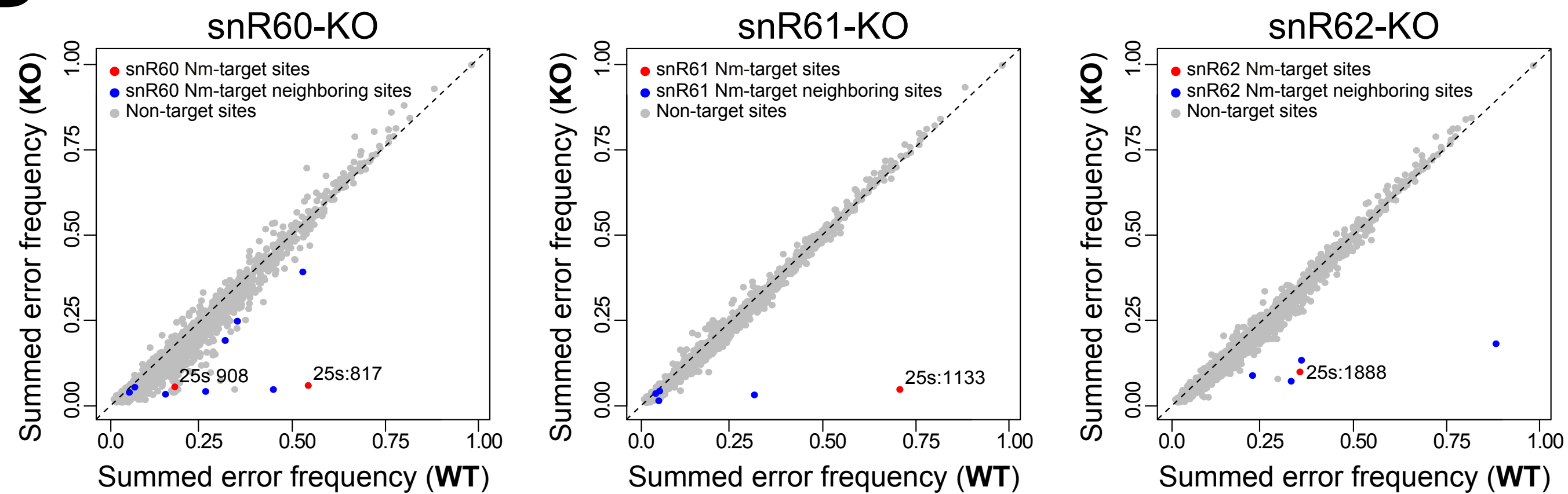
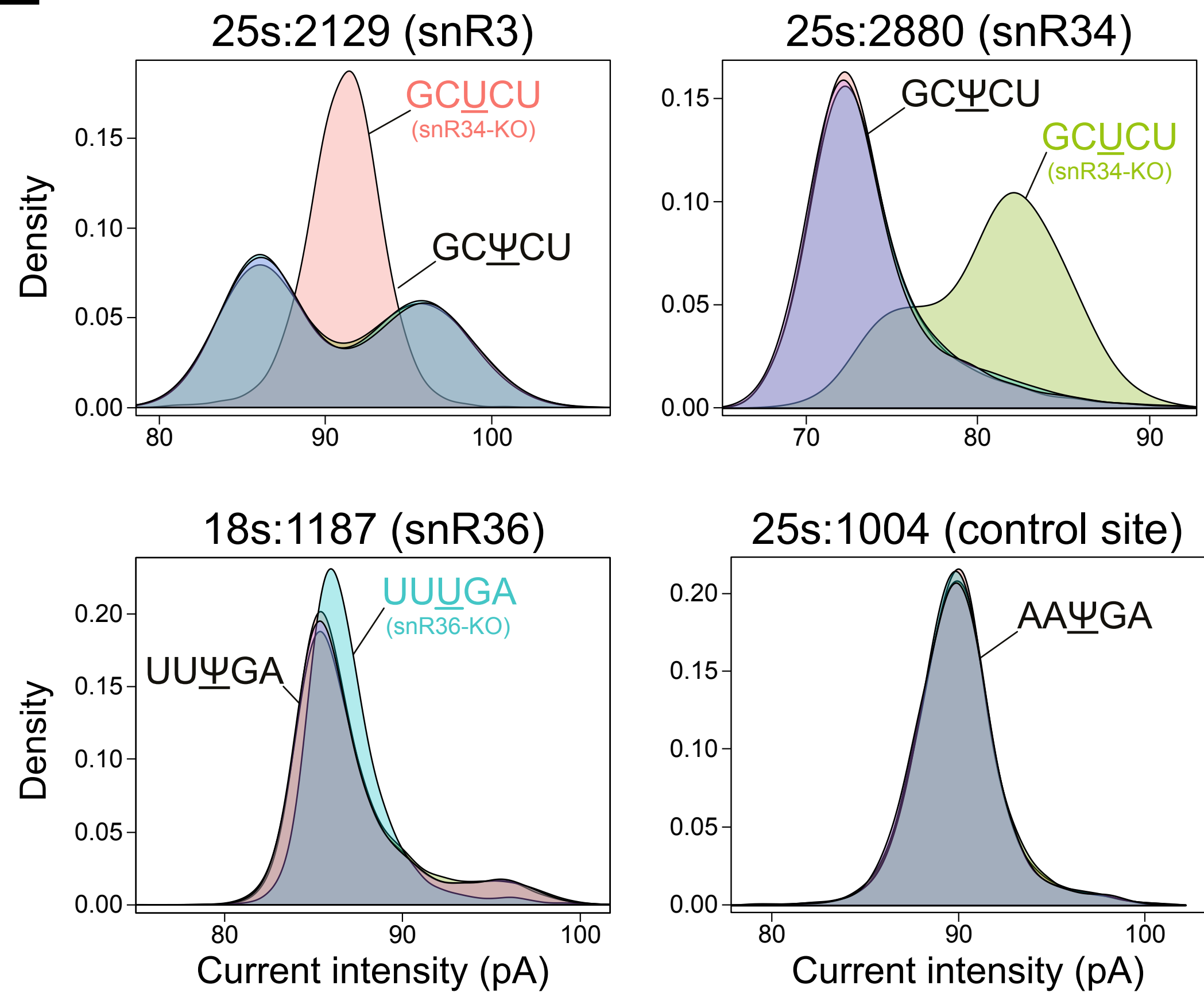
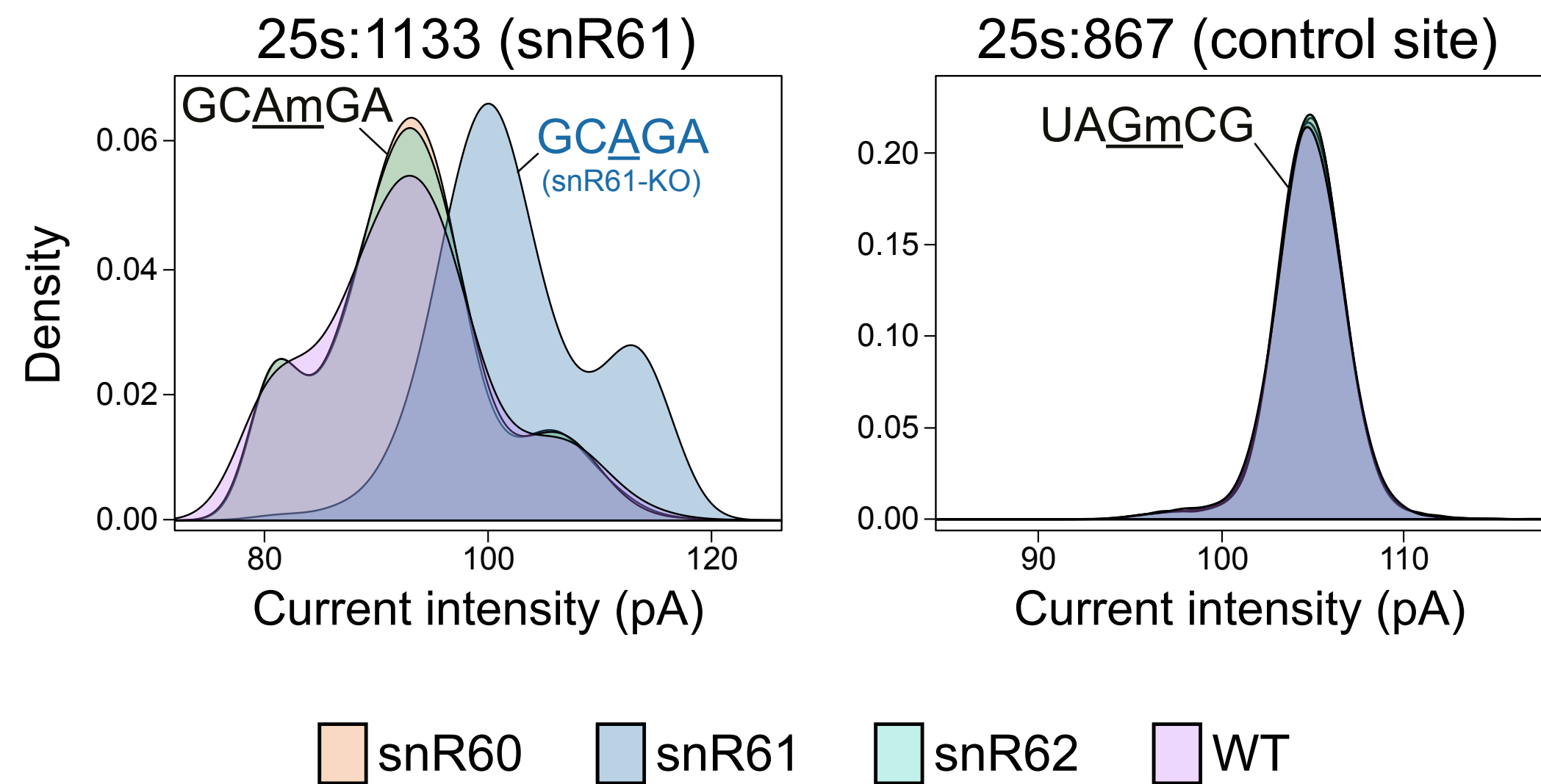


B



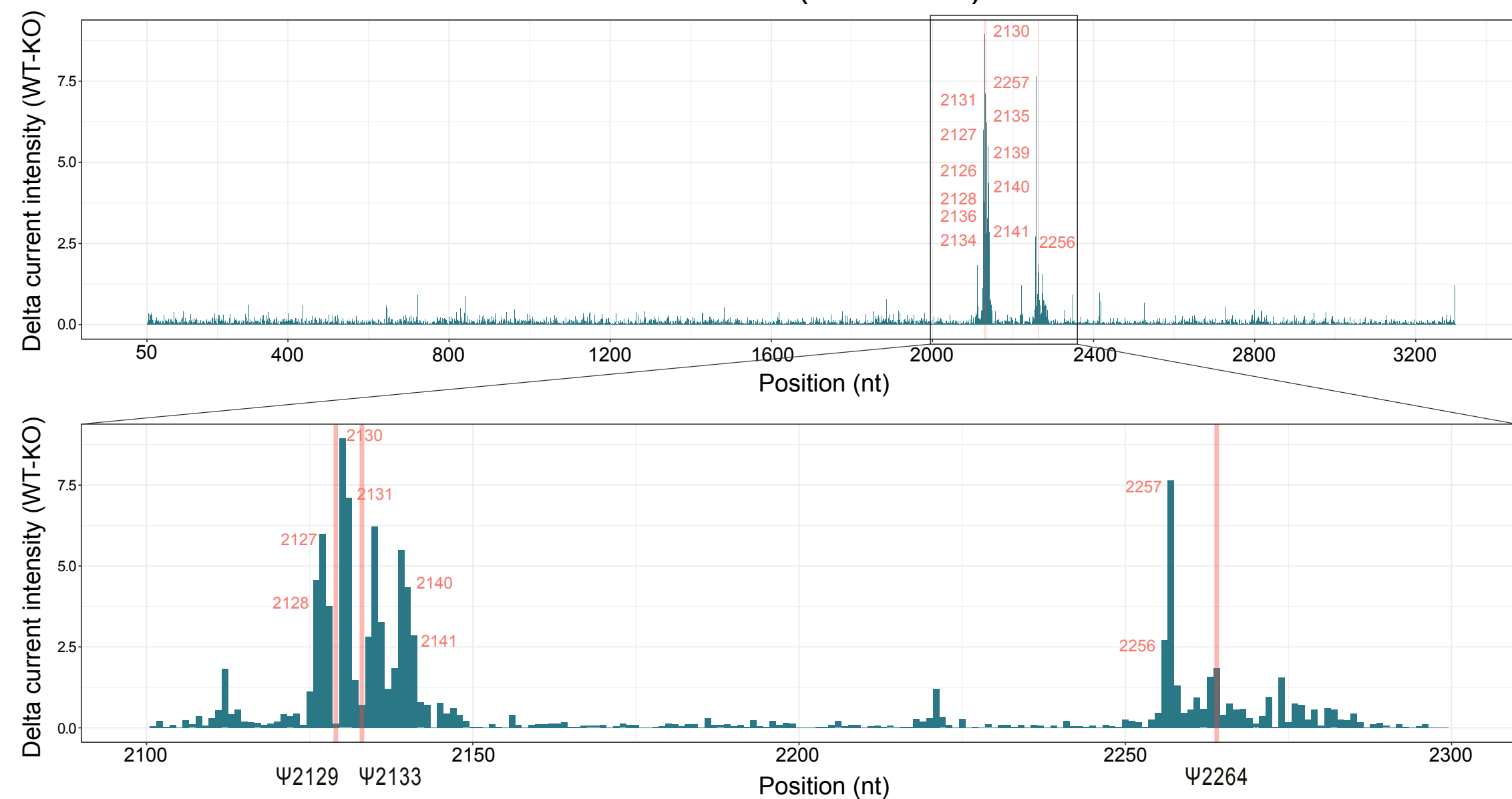
C



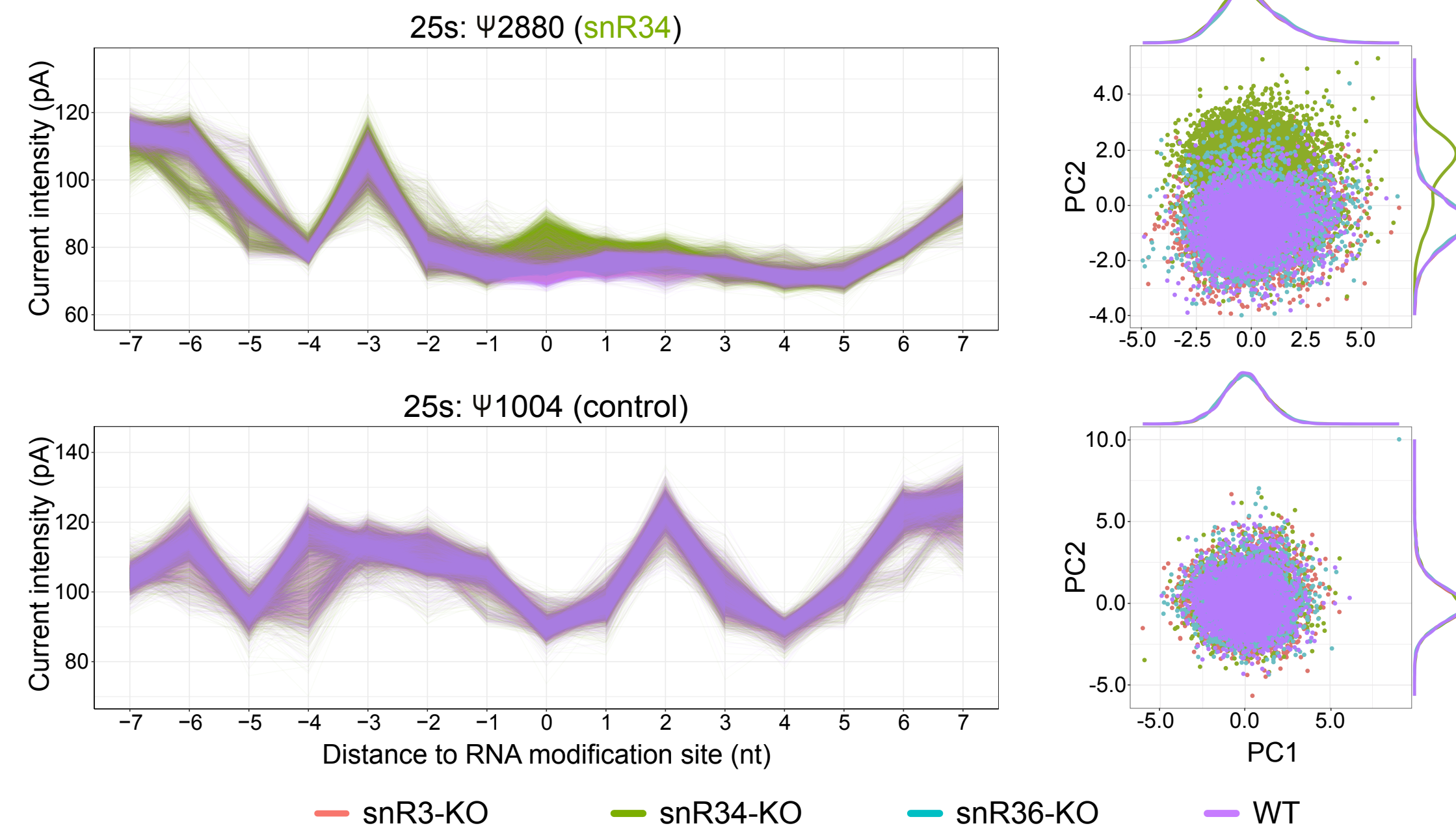
**A****B****C****D****E****F**

A

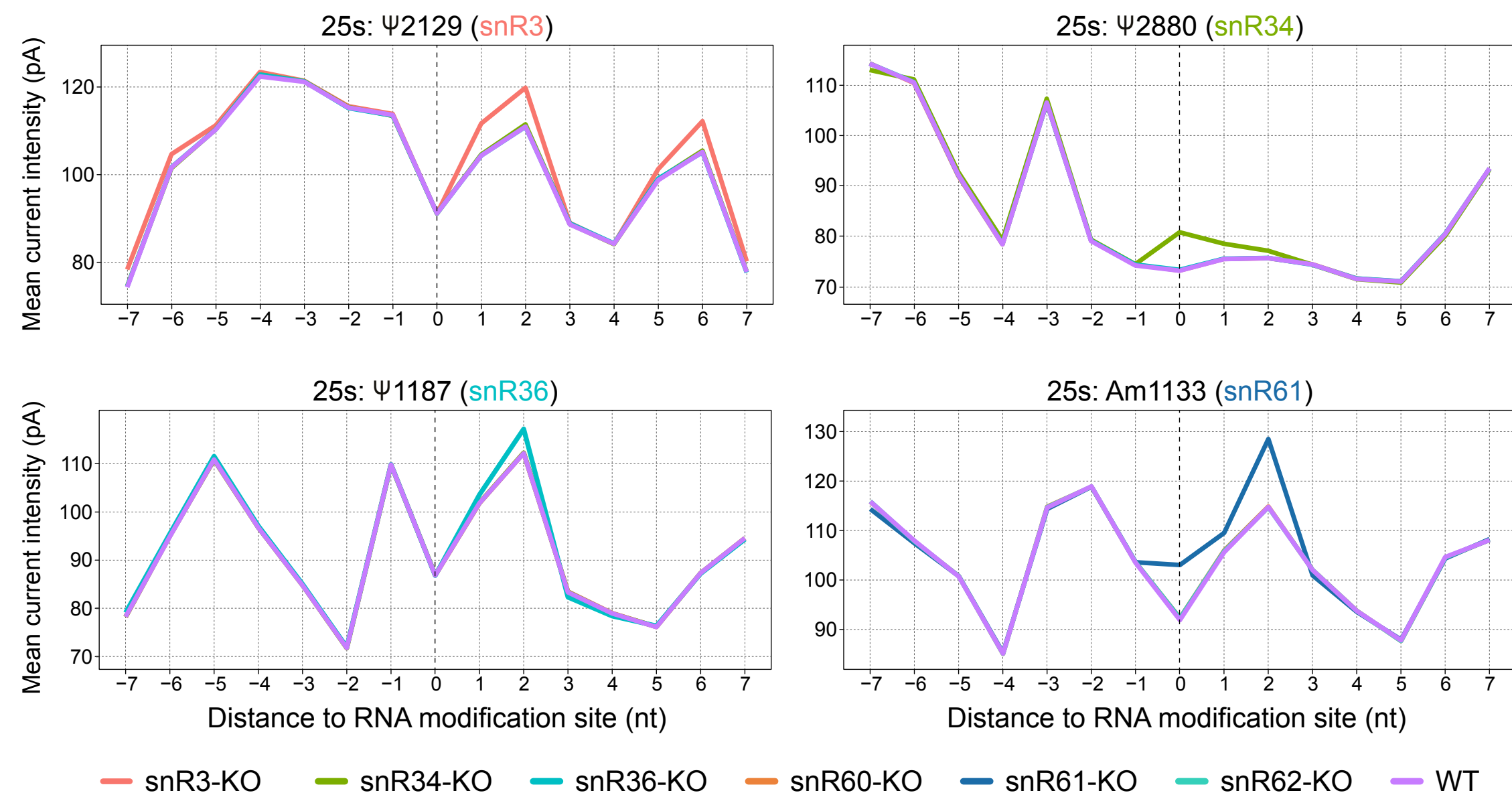
## 25s rRNA (snR3-KO)



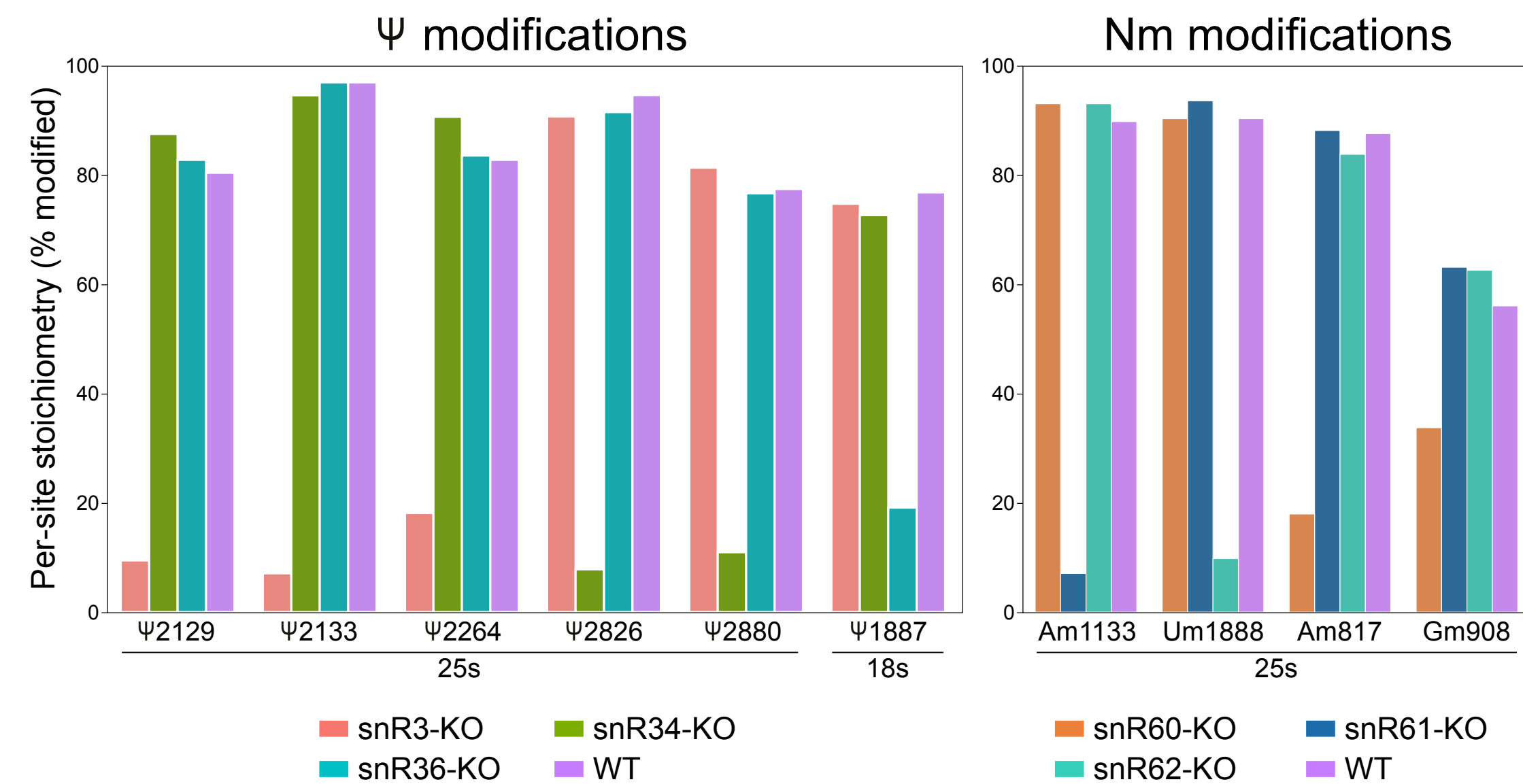
C

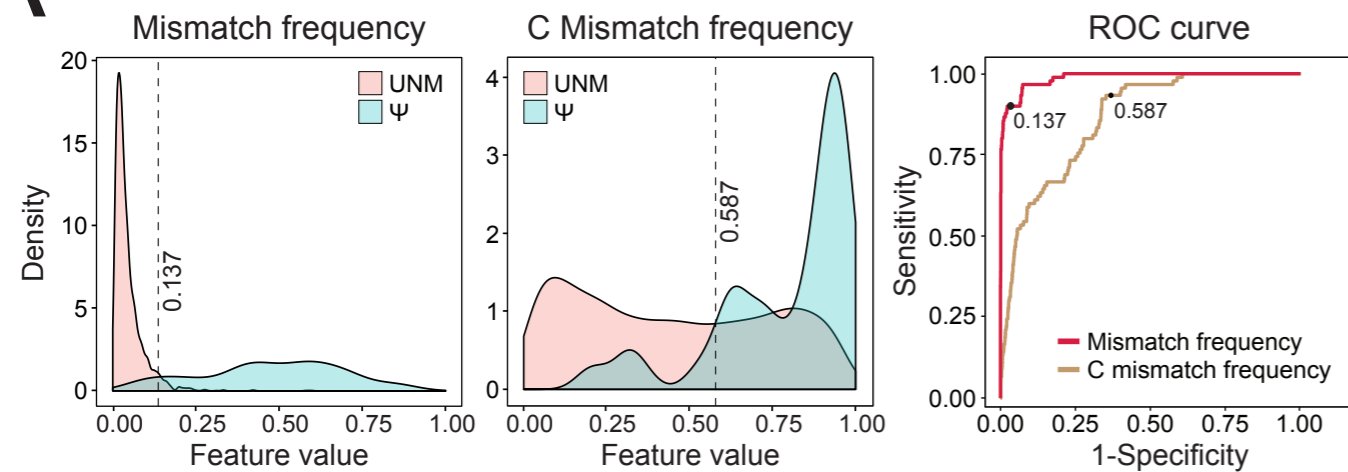
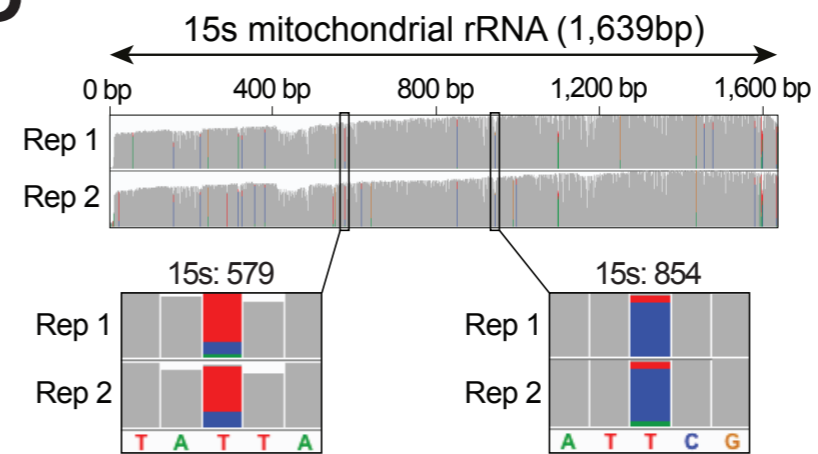
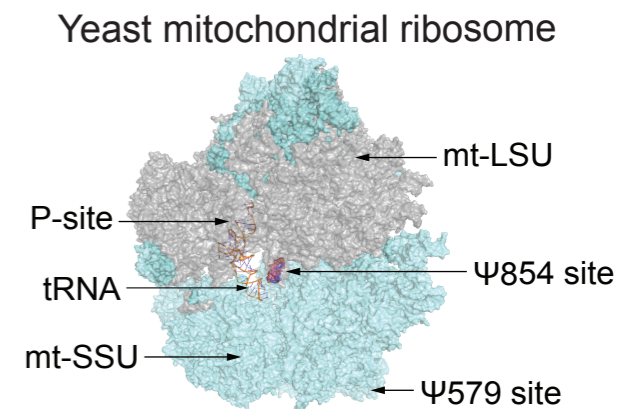
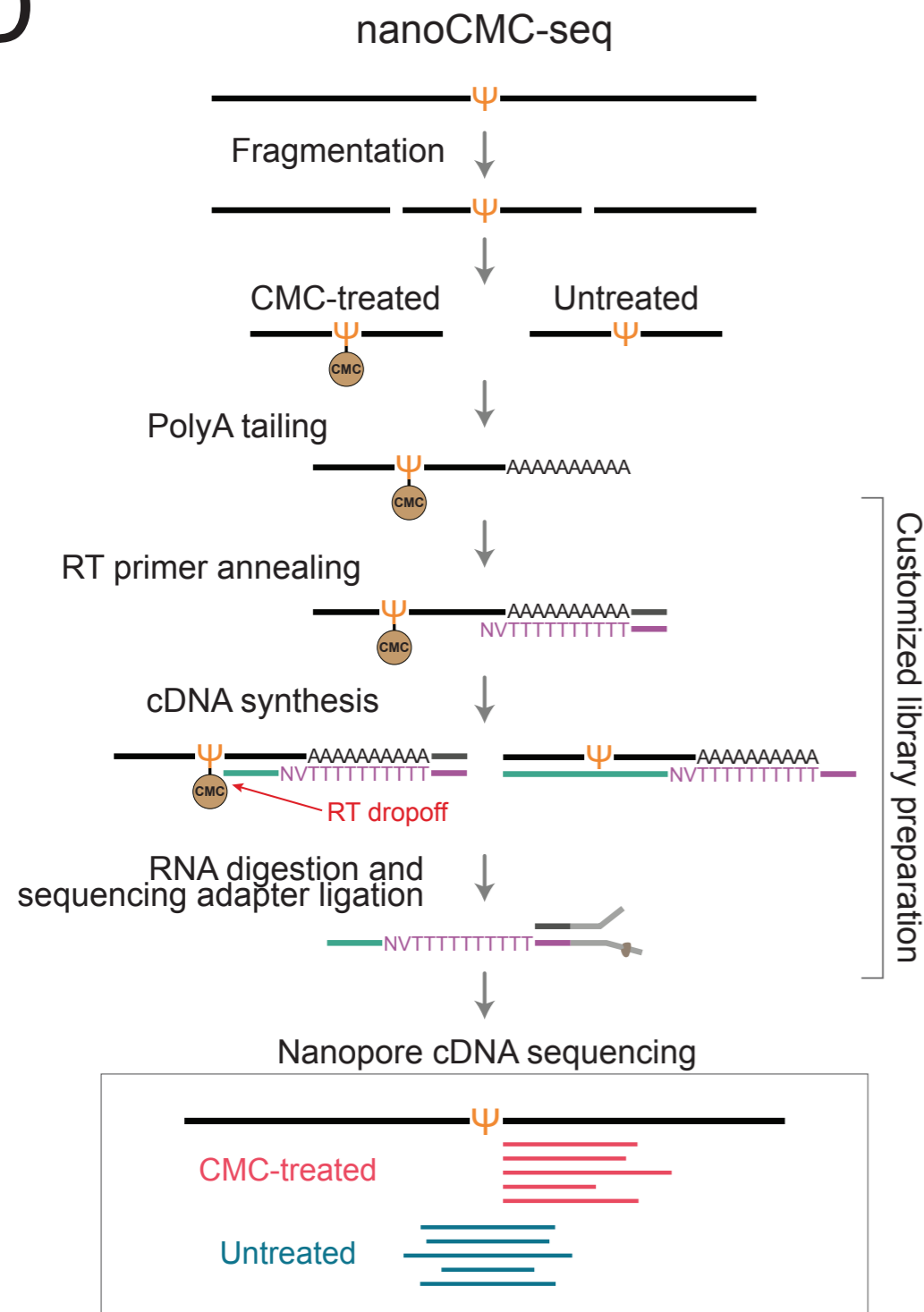
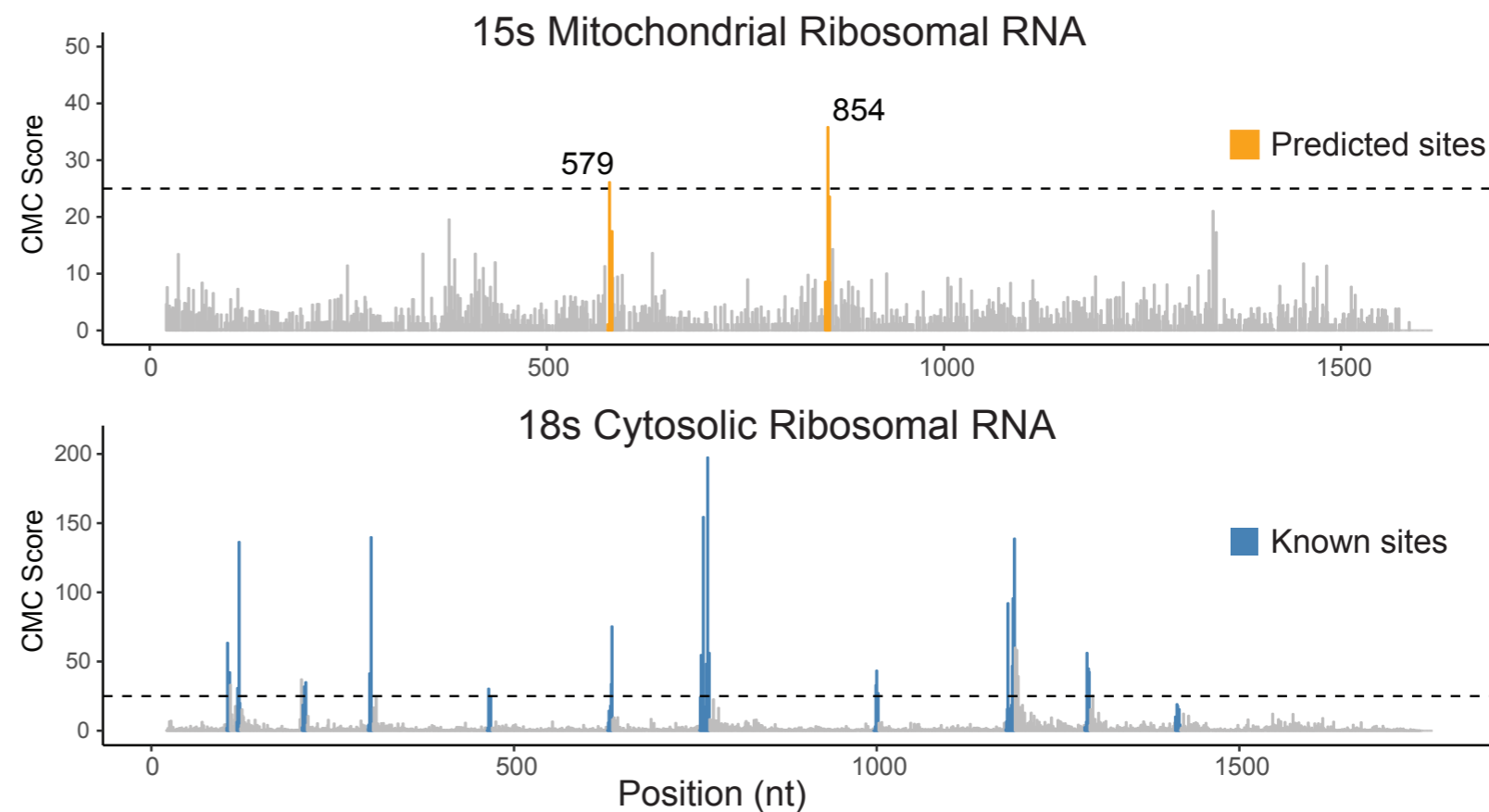
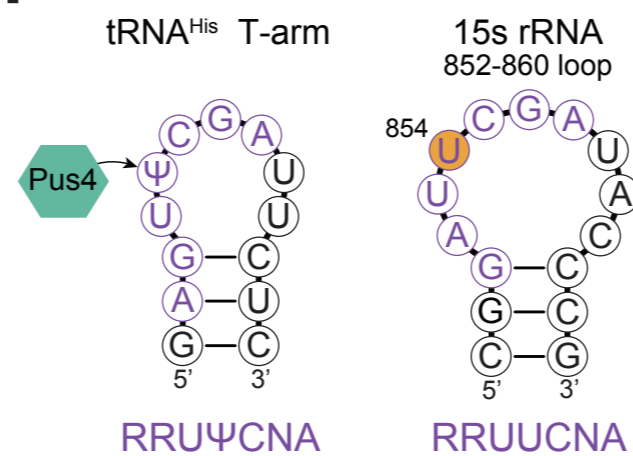
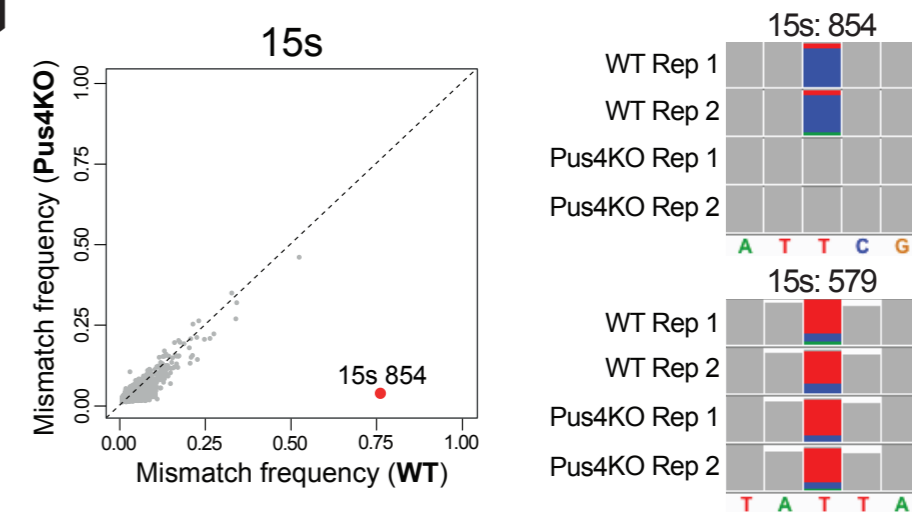


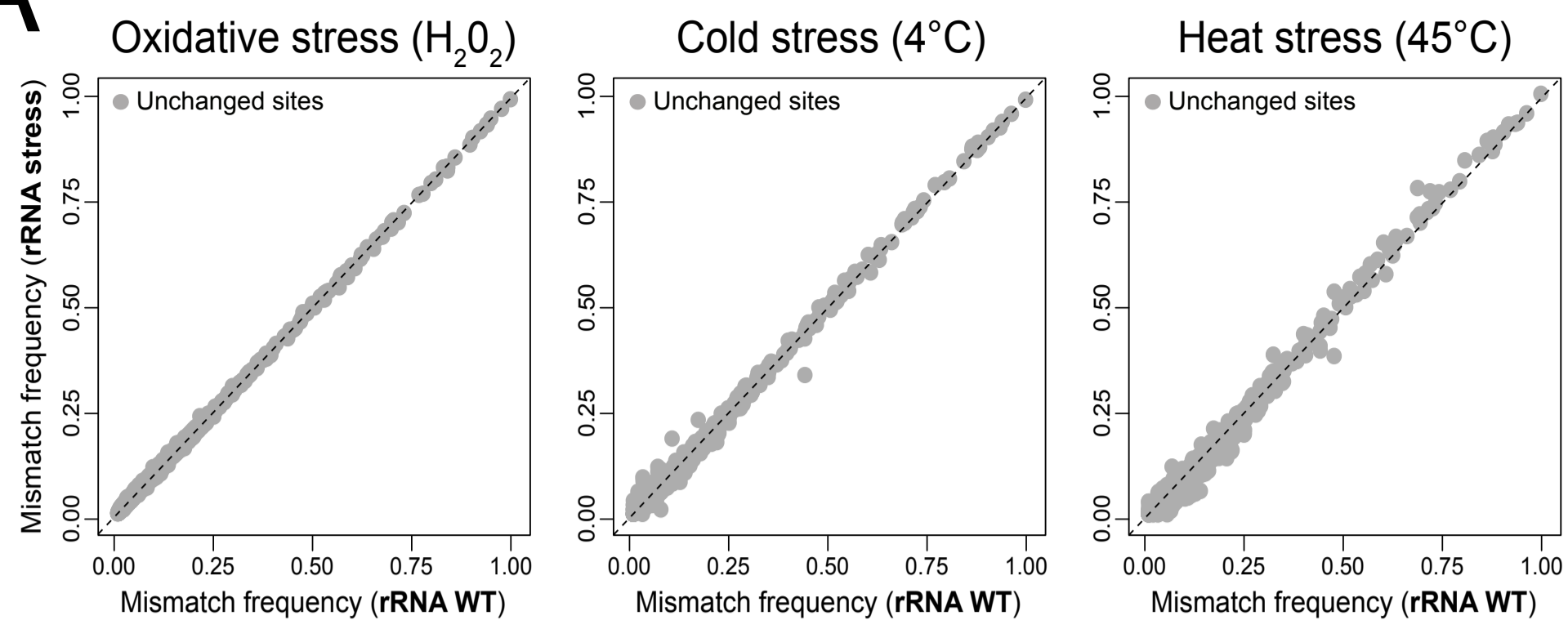
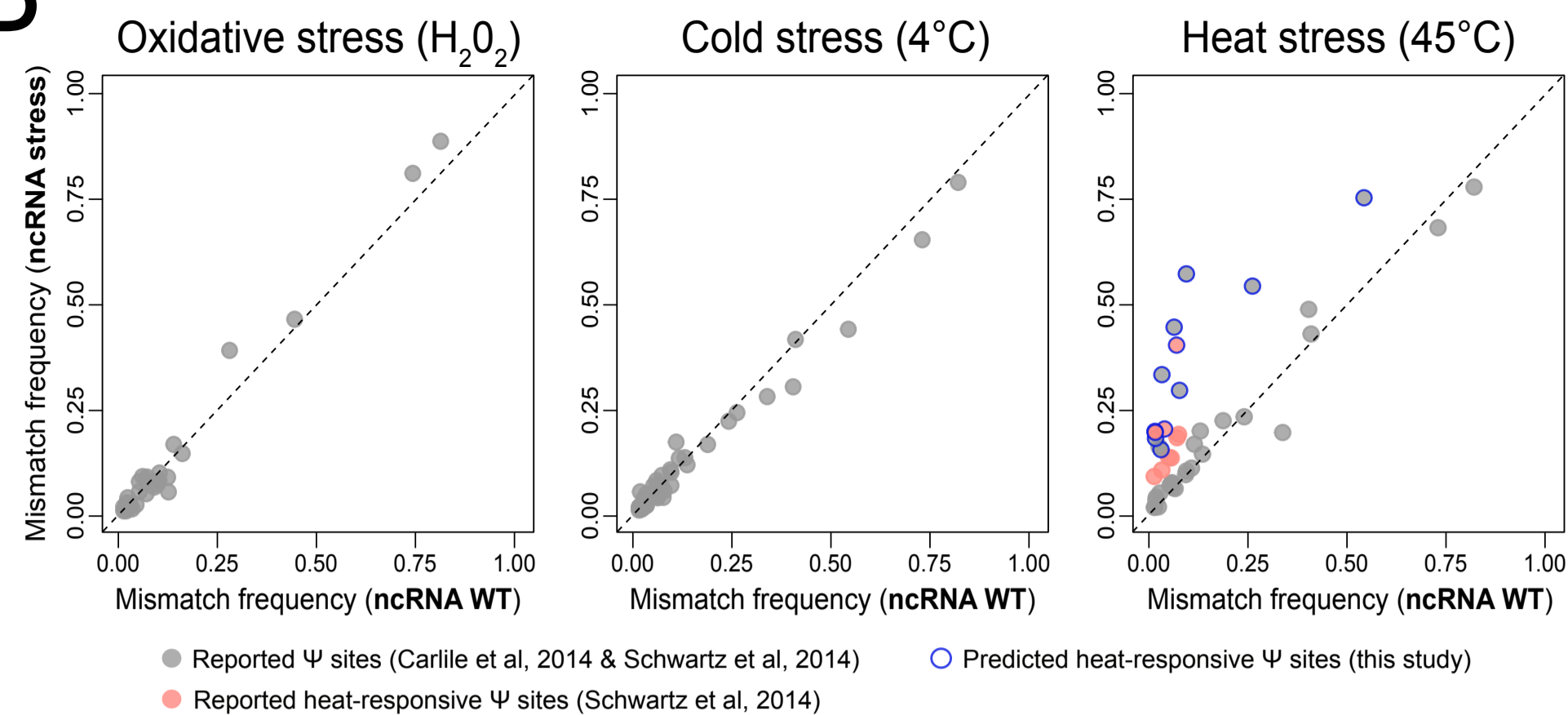
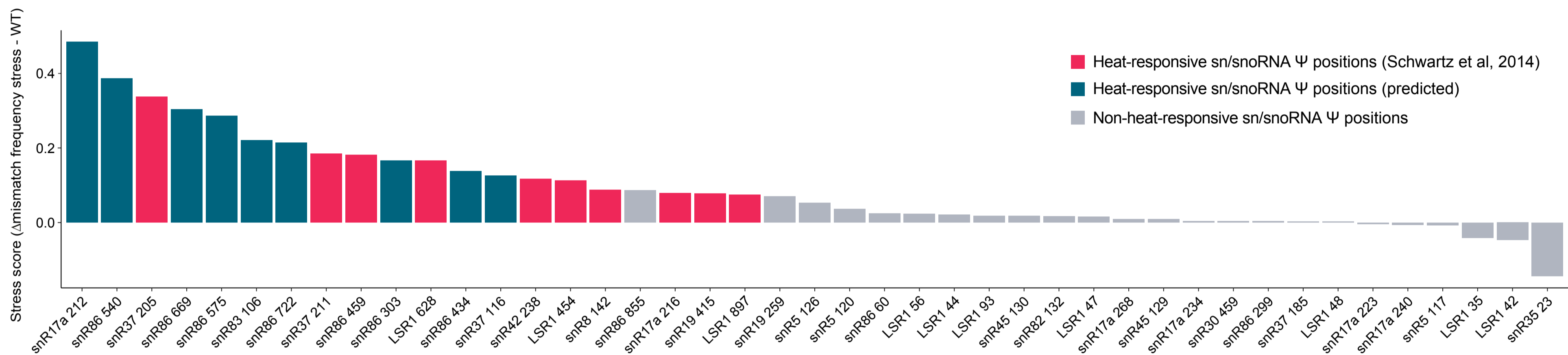
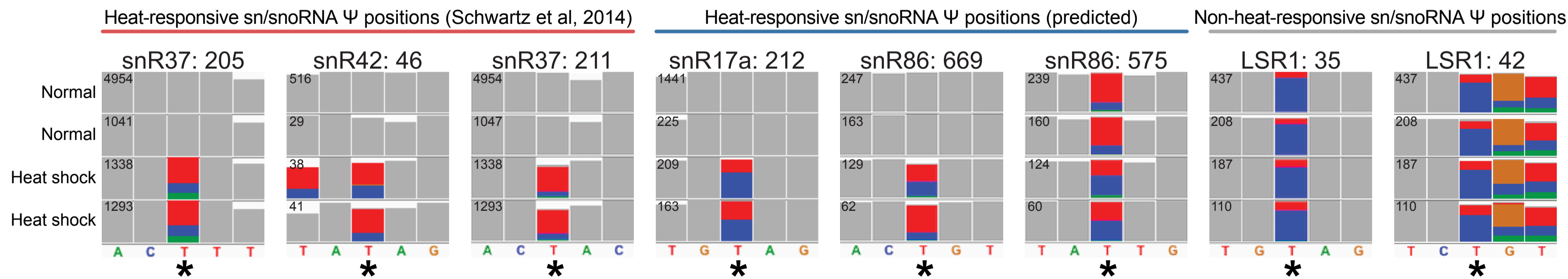
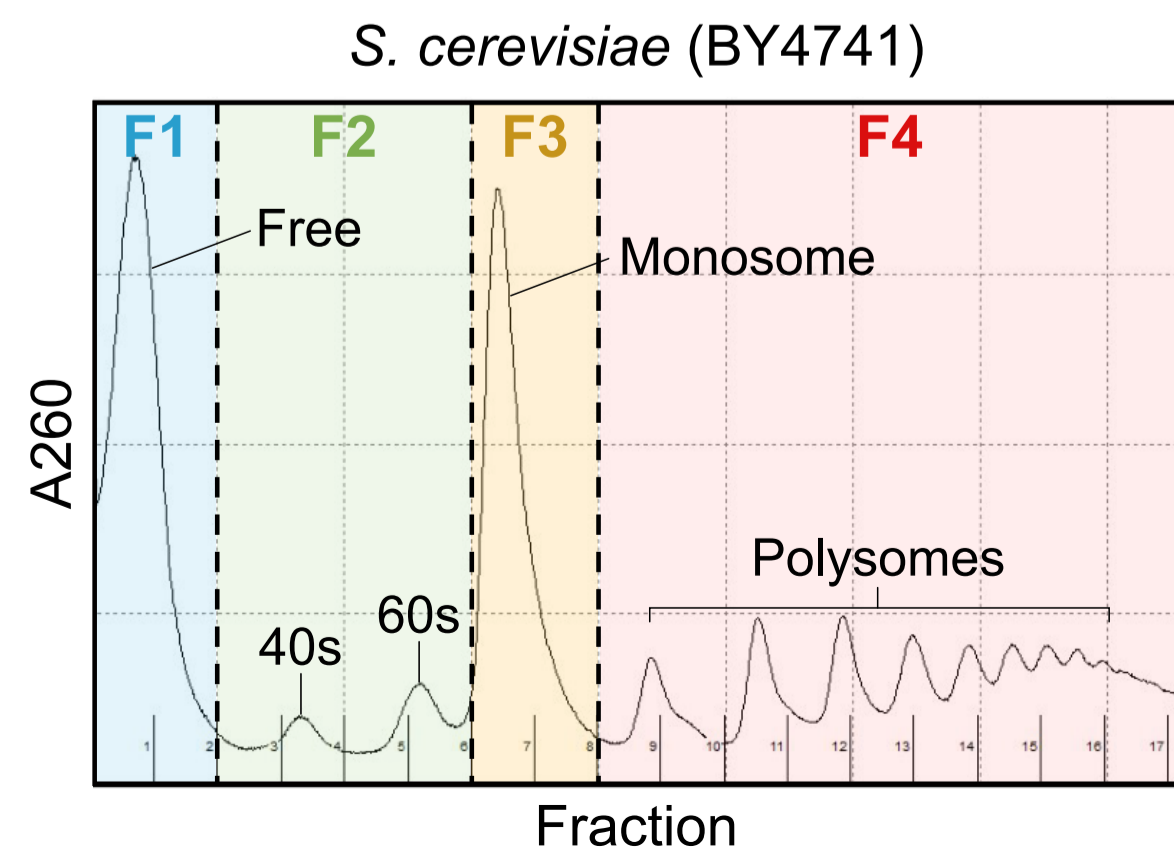
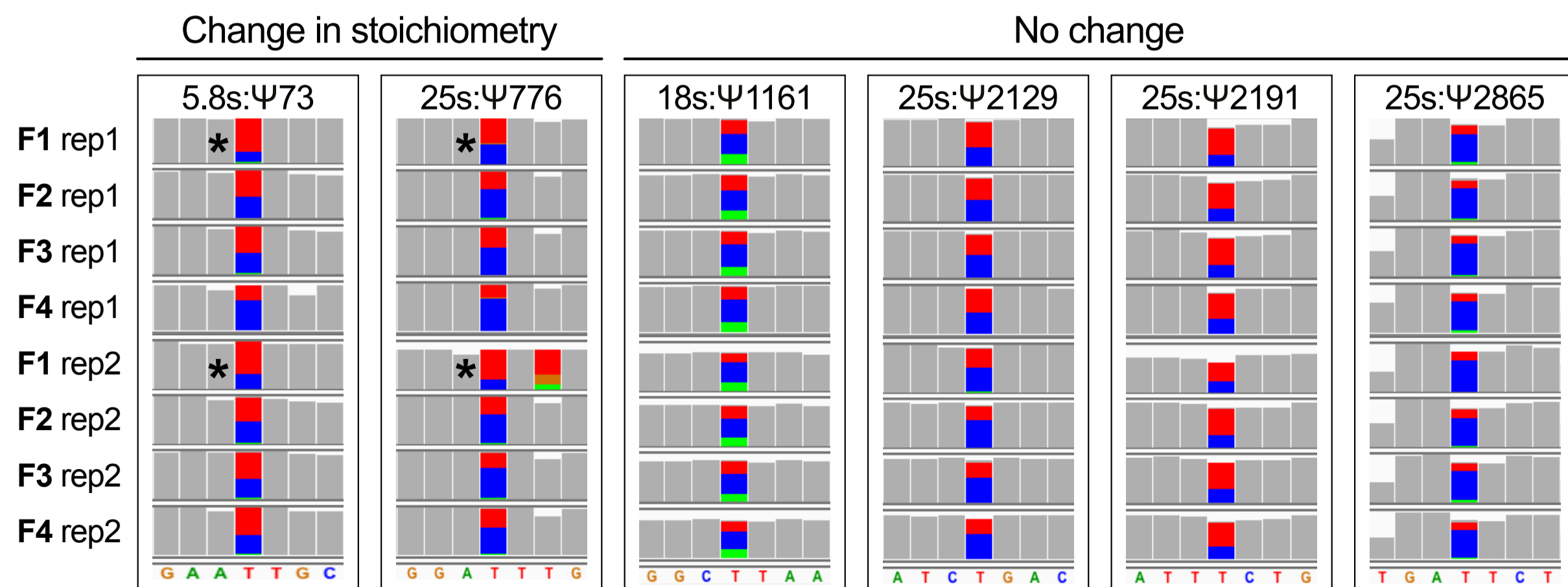
B



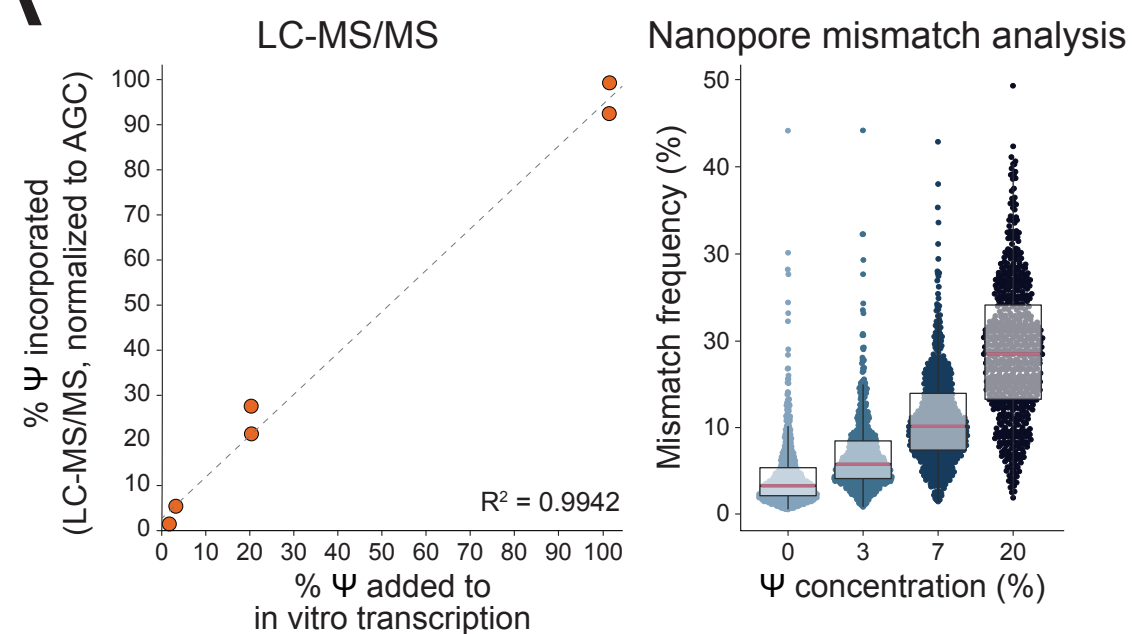
D



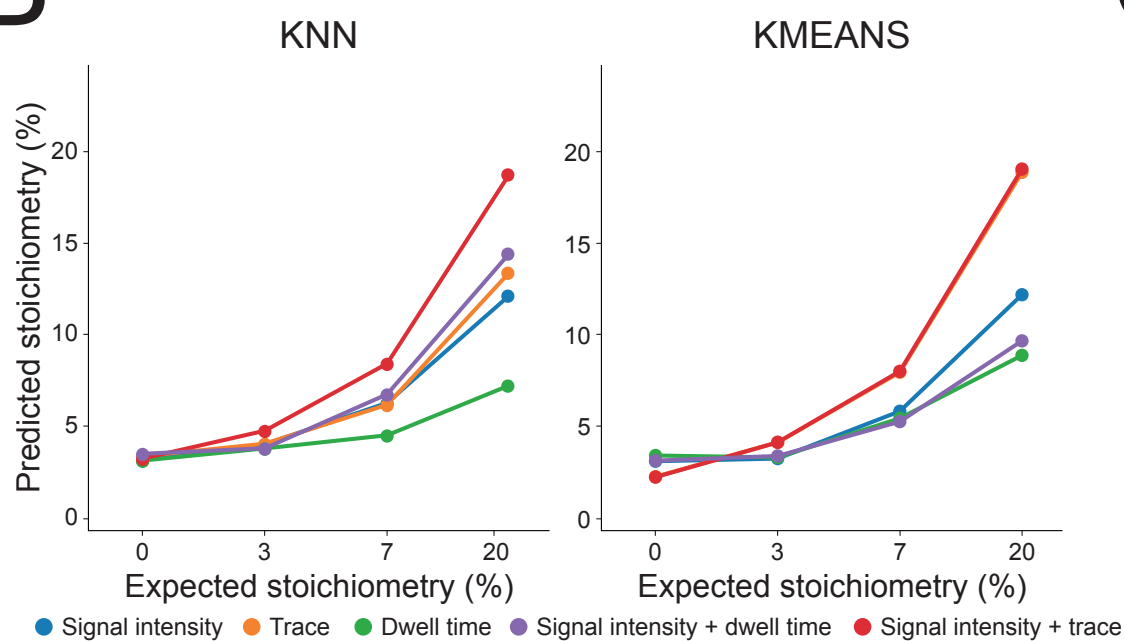
**A****B****C****D****E****F****G**

**A****B****C****D****E****F**

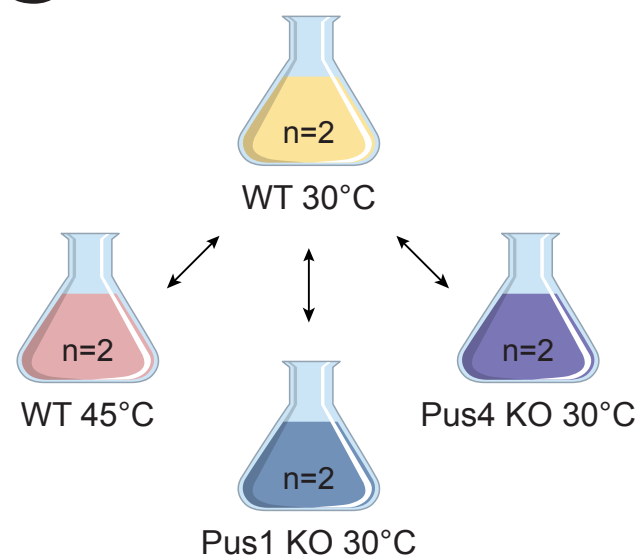
A



B

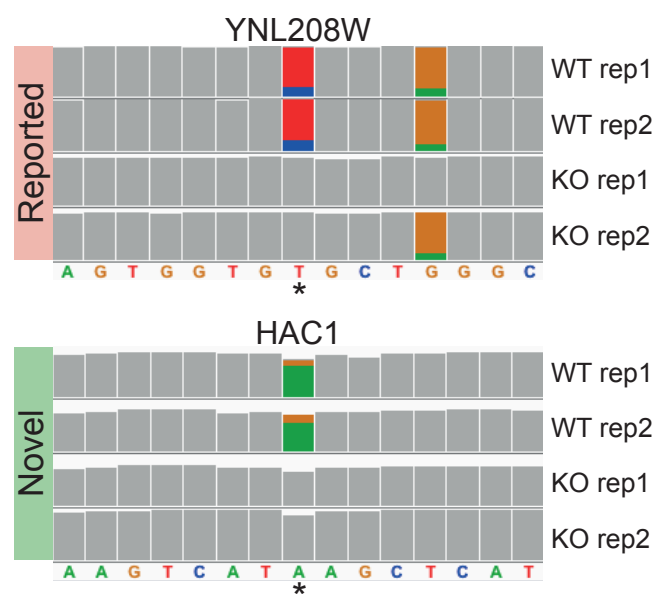
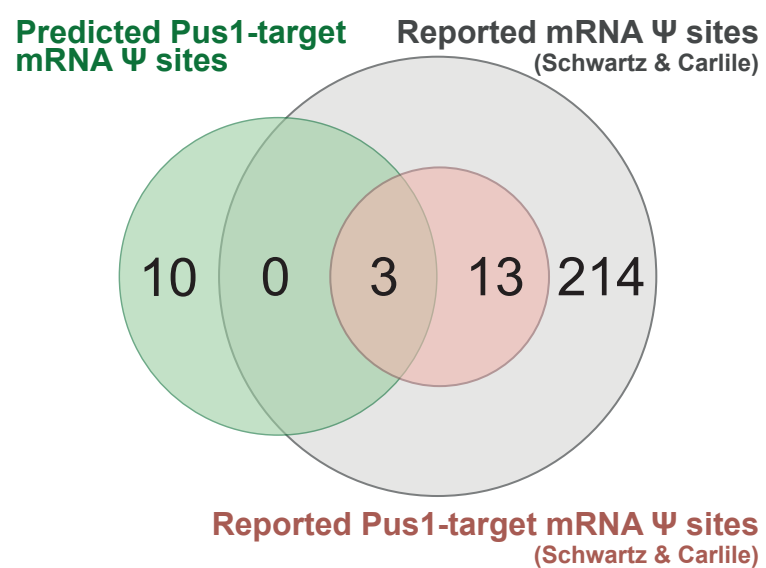


C

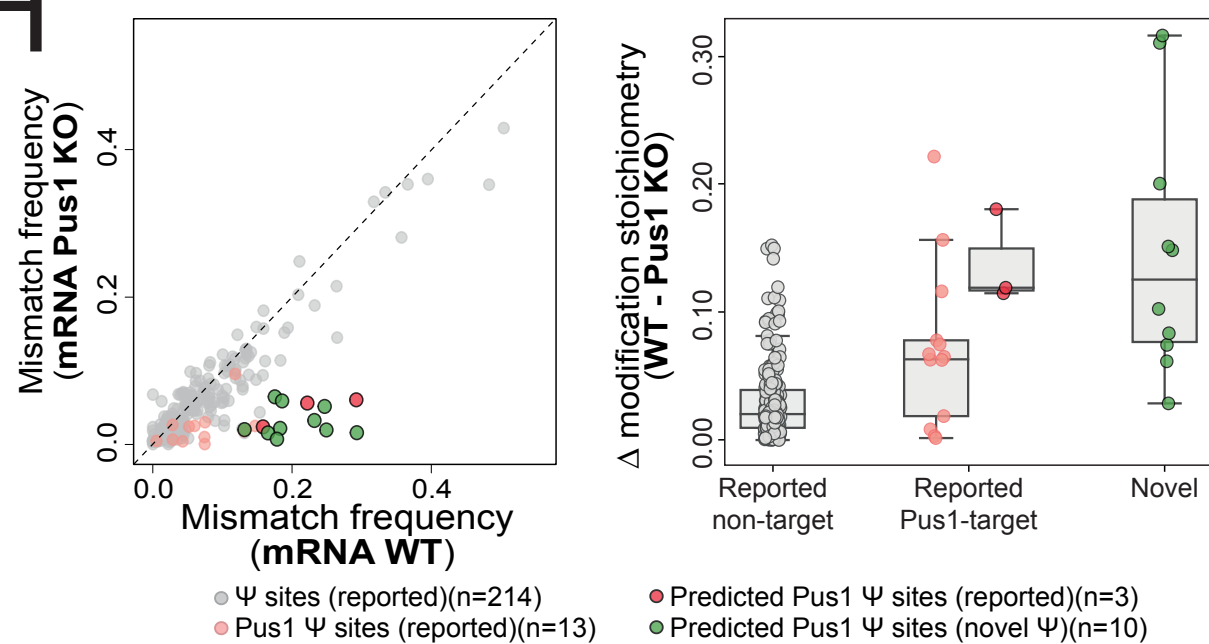


D

mRNA Pus1 KO

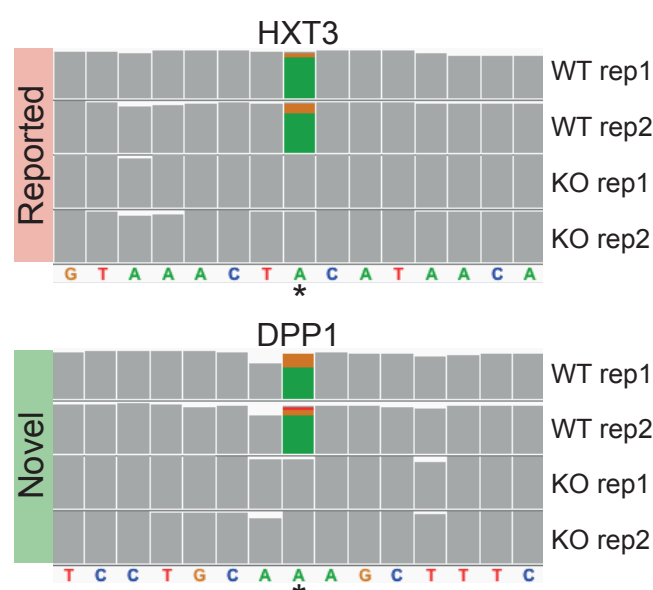
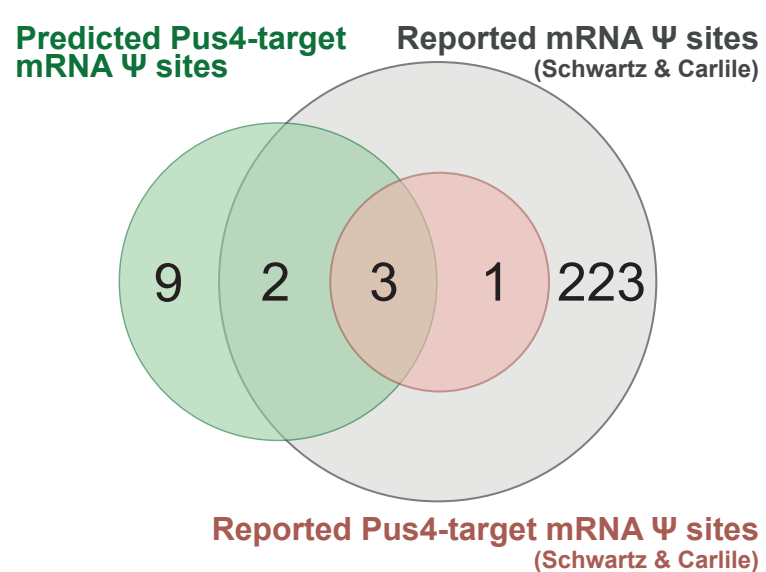


H

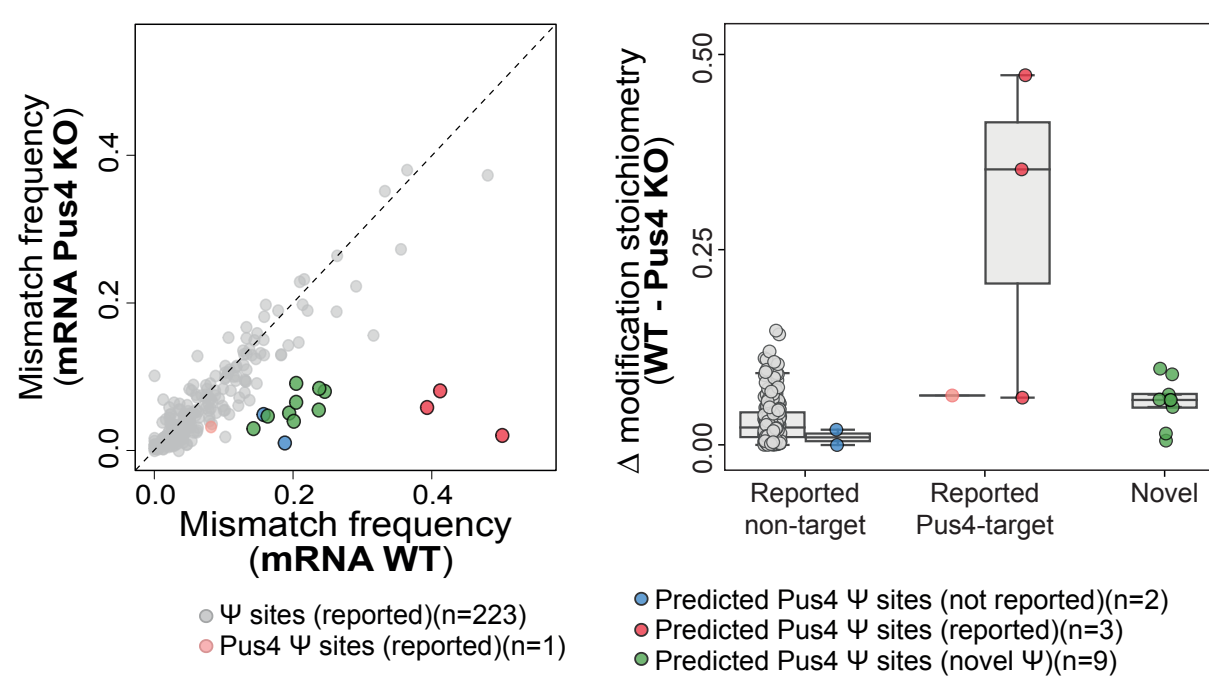


E

mRNA Pus4 KO

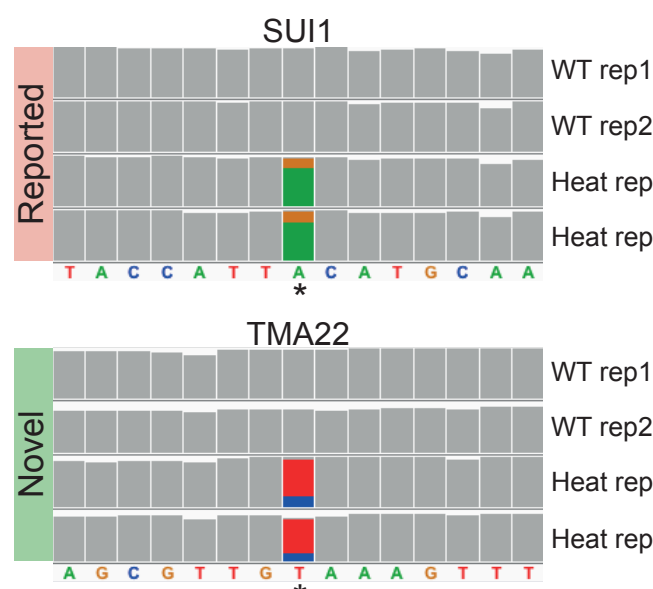
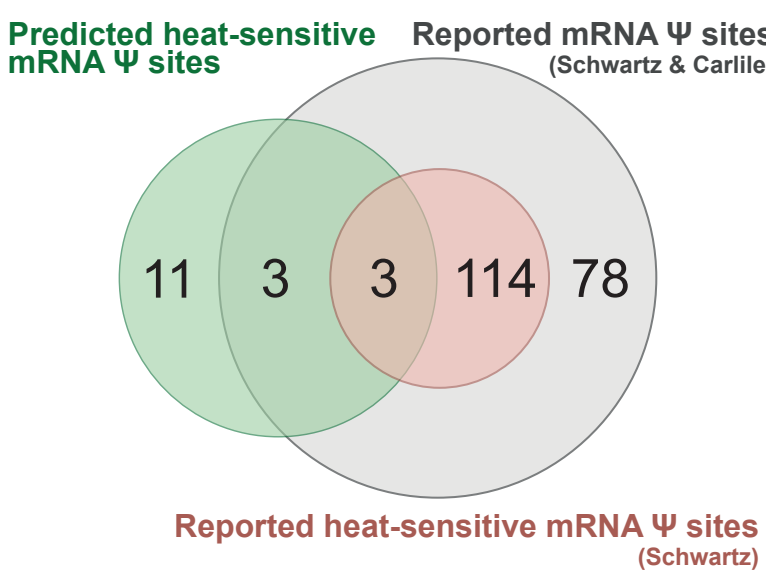


I

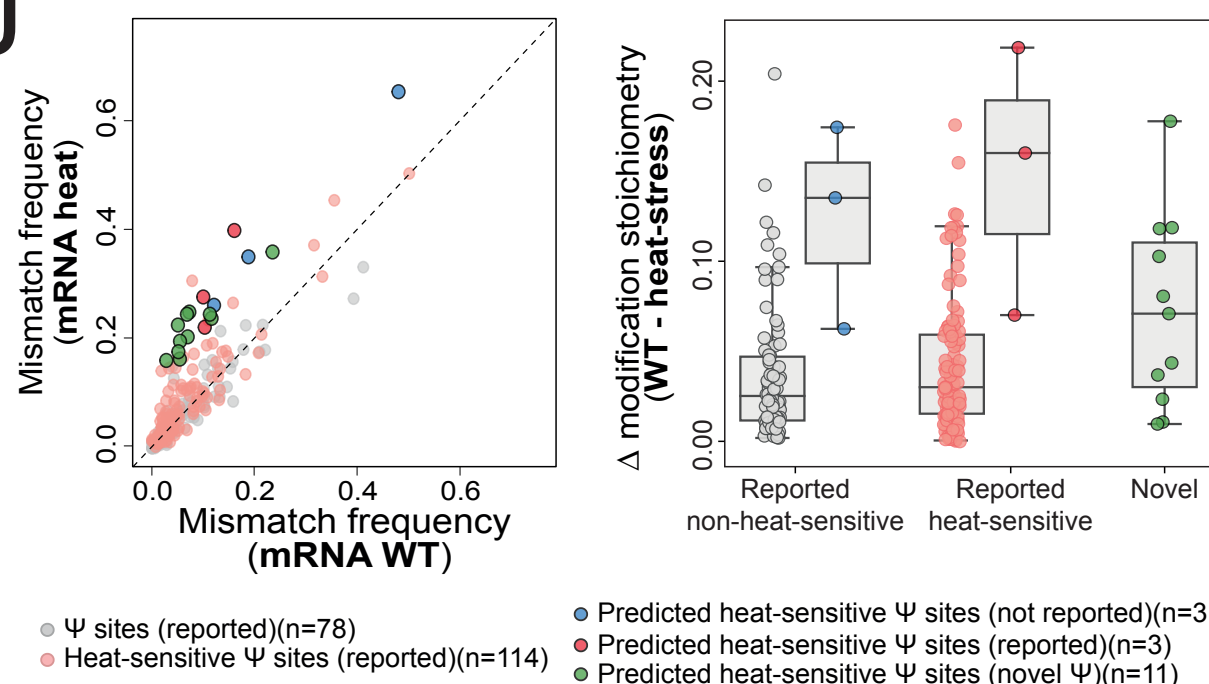


F

mRNA heat-stress

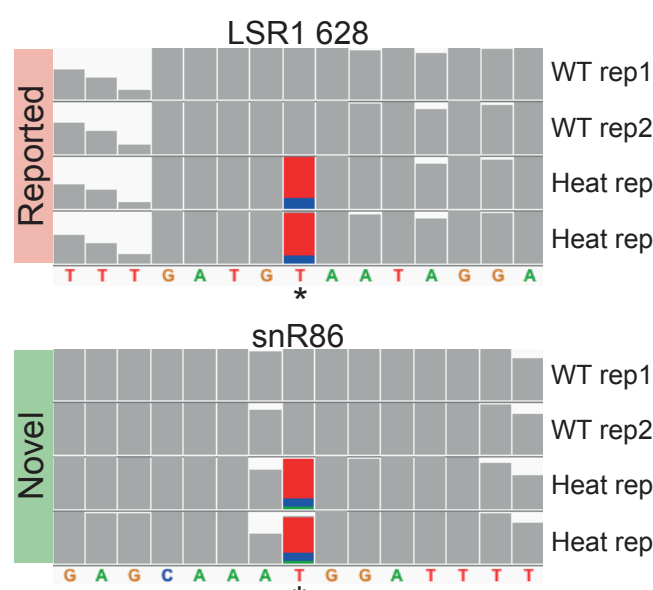
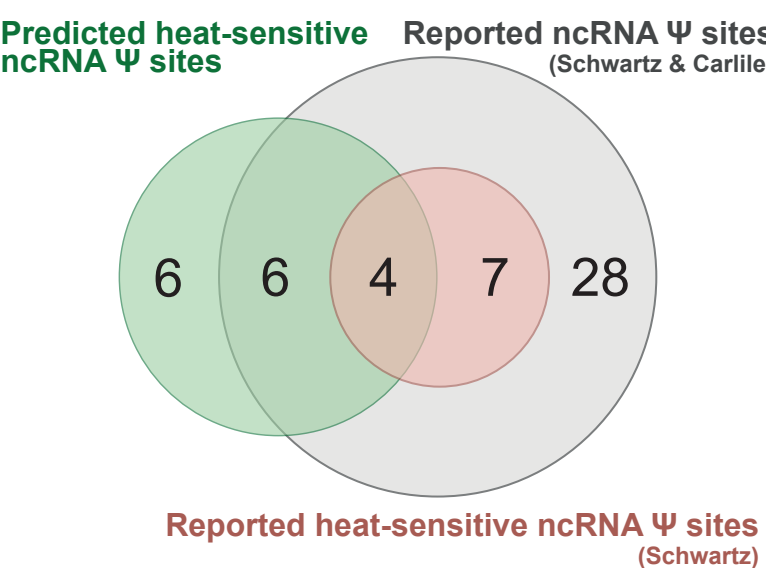


J



G

ncRNA heat-stress



K

