Establishing benchmarks for quantitative mapping of m^{6}A by Nanopore Direct RNA Sequencing

Authors

Denise Ohnezeit*[1], Elene Loliashvili*[2], Gregory Putzel[3], Ruth Verstraten[2,4], Jianheng Liu[5], Luke S. Nicholson[5], Alejandro Pironti[3], Samie R. Jaffrey[5], Daniel P. Depledge[#1,2,4,6], Angus C. Wilson[#1]

Affiliations

[1]Department of Microbiology, New York University School of Medicine, New York, NY, USA

[2]Institute of Virology, Hannover Medical School, Hannover, Germany

[3]Antimicrobial-Resistant Pathogens Program, New York University Grossman School of Medicine, New York, NY, USA

[4]German Center for Infection Research (DZIF), partner site Hannover-Braunschweig, Hannover, Germany

[5]Department of Pharmacology, Weill Medical College, Cornell University, New York, NY, USA

[6]Cluster of Excellence RESIST (EXC 2155), Hannover Medical School, Carl-Neuberg-Straße 1, 30625 Hannover, Germany

* Joint first authors: contributed equally

[#] Joint senior authors: contributed equally

ORCIDs

DO = 0000-0002-2346-2545

EL = 0009-0006-4098-5010

GP = 0000-0002-5211-2608

RV = 0009-0008-6046-6335

JL = 0000-0003-0216-1951

LSN = 0009-0005-8203-7249

AP= 0000-0002-7759-6776

SRJ = 0000-0003-3615-6958

DPD = 0000-0002-4292-0599

ACW = 0000-0002-5016-4164

Abstract

Nanopore direct RNA sequencing (DRS) coupled with Dorado modification-aware basecalling enables mapping of epitranscriptomic modifications including N^6-methyladenosine (m^6A) at the level of individual RNAs. However, a lack of systematic benchmarking continues to raise questions regarding the sensitivity, specificity, and reproducibility of this method. To address this and to establish a best-practice workflow, we evaluated multiple Dorado versions using *in vitro* transcribed RNA and an m^6A methyltransferase inhibitor as specificity controls. We established that stringent filtering is necessary to reduce false-positive calls and found strong concordance at high-stoichiometry sites when compared to an orthogonal m^6A mapping method (GLORI). Further, by applying DRS to primary human fibroblasts and HD10.6 neurons, we uncovered cell type-specific differences in m^6A stoichiometry, indicating a finely tuned epitranscriptomic regulation. Our study thus presents the first systematic comparison of Dorado and GLORI from the same input RNA and expands characterization of the m^6A epitranscriptome to fibroblasts and neurons.

Introduction

Epitranscriptomics describes the occurrence of chemical modifications on RNA that regulate stability, localization, splicing, translation or decay[1]. To date, over 170 RNA modifications are known to occur across all classes of RNA including eukaryotic mRNA[2]. Of these, at least thirteen different chemical modifications have been characterized in mRNAs, comprising both cap-proximal and internal modifications[3]. By contrast, current evidence indicates that lncRNAs harbor a more limited repertoire of modifications, primarily N^6-methyladenosine (m^6A), 5-methylcytosine (m^5C), and N7-methylguanosine (m^7G)[4]. One of the best studied and most abundant internal mRNA modifications is m^6A. Functional studies of m^6A methylases, demethylases and m^6A binding proteins, as well as sequencing-based mapping methods have increased

53 our understanding of m^6A as a reversible and dynamic modification that fine-tunes nearly all aspects of
54 mRNA metabolism[1,5-8].

55 In addition to m^6A, other widespread and functionally important marks include m^5C, which is less
56 abundant but influences mRNA export and stability, pseudouridine (Ψ), which contributes to mRNA
57 stability, translation and splicing, and adenosine-to-inosine (A to I) editing, which is involved in splicing,
58 stability and recoding events that contribute to protein diversity[9].

59 Techniques based on short-read (Illumina) sequencing have advanced in recent years to achieve single-
60 nucleotide resolution and quantitative mapping of RNA modifications. The approaches that have emerged
61 to map m^6A can be broadly classified as antibody-based, enzyme-assisted or chemical conversion
62 methods[10].

63 Early antibody-based techniques[11-13] provided the first transcriptome-wide insights into m^6A architecture,
64 albeit at limited resolution. Enzyme-assisted methods[14,15] have improved precision but can introduce
65 enzymatic biases. These biases are avoided with GLORI-Sequencing (GLORI), which selectively deaminates
66 unmethylated adenosines, enabling theoretically unbiased, quantitative, single-nucleotide mapping of
67 m^6A[16].

68 Despite advances, the broad utility of these approaches is limited by a lack of isoform specificity, as well
69 as biases associated with amplification, specificity of antibodies or variable adenosine conversion rates[17].
70 Oxford Nanopore Technologies (ONT) direct RNA sequencing (DRS) offers an alternative to these methods
71 by analyzing native RNA molecules that have retained their modifications[18]. As single RNAs pass through
72 nanopores, signal changes are interpreted to determine canonical and modified residues at single-
73 nucleotide resolution on individual transcripts[19,20]. For the now discontinued DRS SQK-RNA002 chemistry
74 used with R9 flowcells, multiple research groups developed algorithms for detecting RNA modifications.
75 These methods can be divided into error rate-based and current signal-based[21]. While the recent release
76 of the DRS SQK-RNA004 chemistry and RNA-specific flowcells reduced the utility of many of these tools,
77 introduction of a modification-aware basecaller (Dorado) provides an all-in-one toolkit for basecalling,
78 RNA modification detection, and poly(A) tail length estimation. Early versions of Dorado (e.g. v0.5.0 and
79 v0.6.0) enabled m^6A (DRACH-context only) detection, while version v0.7.0 enabled detection of Ψ and all-
80 context m^6A. More recent versions (v0.8.0 and v0.9.0) allow both DRACH- and all-context m^6A detection,
81 together with Ψ, inosine and m^5C[22].

82 In a recent overview of this emerging technology, Cruciani and Novoa highlighted the importance of
83 rigorous evaluation of the limitations of Dorado and advocated forcefully for the establishment of best
84 practices to ensure consistency and accuracy going forward[23]. To date, only a handful of studies of this
85 kind have been published. A series of studies from Novoa and colleagues showed that Dorado performs
86 robustly on synthetic modified and unmodified RNA molecules but also found that overlaps in the
87 identification of specific modifications and false-positive calls remain a challenge but can be mitigated by
88 the inclusion of *in vitro* transcribed (IVT) controls[17,23]. Independently, Esfahani and colleagues used poly(A)
89 RNA from a B-lymphocyte cell line to compare SQK-RNA002 and SQK-RNA004-derived DRS data. While
90 they concluded that Dorado outperformed other RNA modification detection tools, IVT controls were
91 necessary to reduce false-positive calls[24]. Recent benchmarking efforts comparing Dorado and m6Anet on
92 SQK-RNA004 chemistry have shown that both tools achieve high recall and strong stoichiometry
93 correlations at DRACH motifs but exhibit higher false discovery rates on IVT controls and lower recalls of
94 low-stoichiometry genomic m^6A sites. However, these evaluations relied on previously published GLORI
95 and eTAM-Seq datasets, rather than matched input material[17,22]. By directly comparing m^6A detection by
96 DRS coupled with Dorado and GLORI on the same RNA inputs, our study overcomes this limitation.
97 Specifically, we systemically benchmark multiple versions of Dorado using poly(A) RNA from primary
98 (normal) human dermal fibroblasts (NHDFs) and functional human neurons produced through terminal
99 differentiation of HD10.6 cells[25], along with IVT RNA. We included treatments with STM2457, an enzymatic
100 inhibitor of the primary m^6A methyltransferase METTL3[26], to robustly assess false-positive m^6A detection
101 rates and also compared the DRS outputs with those obtained from GLORI[27] using the same RNA as input.
102 While our analysis is centered on the sensitivity and specificity of m^6A detection in DRS datasets, we also
103 benchmark the ability of Dorado to reliably detect m^5C, Ψ, and inosine. Together, our data provides a
104 robust workflow for DRS-based modification detection that reliably captures transcriptome-wide, isoform-
105 level, site-specific stoichiometries of RNA modifications.

106

107 ## Results

108 **Systematic benchmarking of ONT's Dorado across human cell types.**

109 To evaluate the performance of ONT's basecalling tool Dorado across different human cell types, we
110 selected primary NHDFs and differentiated HD10.6 cells as a neuron-like model. Prior to isolation of total
111 RNA, cells were treated for 48h with a small molecule inhibitor of METTL3 (STM2457) or with vehicle

112 (DMSO). We then isolated poly(A) RNA and subjected the same input material to both DRS and GLORI. The
113 same material was also used to generate IVT RNA (**Fig. 1A**). Dorado basecalling using version v.0.9.0
114 resulted in over 13M reads using poly(A) RNA from NHDFs treated with DMSO or STM2457, while we
115 obtained over 5M and 10M reads from DMSO- and STM2457-treated HD10.6 cells, respectively. The rate
116 of primary alignment to the human genome (hg38) was over 98% for all the sequenced samples. A detailed
117 overview of read statistics generated by DRS can be found in **Supplementary Table 1**.

118 To systematically benchmark Dorado, we performed basecalling using versions v0.5.0 through v0.9.0.
119 Reads were aligned to the human genome and transcriptome to obtain comprehensive information on
120 isoform-level modification detection. Further, we explored different aspects of detected modifications
121 using the outputs generated by ONT's Modkit tool[28]. We assessed multiple filtering strategies using base
122 and modification probabilities, as well as stoichiometry distributions, and evaluated them based on the
123 IVT data and sensitivity to METTL3 inhibition. By filtering out false-positive calls using multiple control
124 layers, we determined high-confidence m^6A sites (**Fig. 1B**). Finally, we compared these high-confidence
125 predictions with sites identified by GLORI performed on the same input RNA. Using these results, we
126 developed a benchmarked pipeline that filters m^6A sites predicted by Dorado to minimize false-positive
127 detection without impacting sensitivity (**Fig. 1C**).

128 **Stringent filtering of modification probabilities removes false positives**

129 During modification-aware basecalling, each nucleotide is assigned a basecall probability score ranging
130 from 0 to 1. Similarly, individual modified nucleotides are assigned a modification probability score in the
131 same range. To establish a baseline for false-positive m^6A predictions, we plotted basecall and modification
132 probability distributions for an IVT poly(A) RNA dataset processed using Dorado v0.9.0 and aligned to
133 human genome (**Fig.2**). We binned basecall and modification probability values for A and m^6A (all-context
134 and DRACH) and compared this data from DMSO- and STM2457-treated NHDF datasets.

135 We observed that 75-90% of A basecalls (in all datasets) have probability scores above 0.8 **(Fig. S1A)**. Since
136 all A bases are considered during all-context m^6A calling, the fraction of m^6A sites is substantially smaller
137 when compared to DRACH-context m^6A sites (**Fig.2A,B, Fig.S1B,C**). The modification probability
138 distributions are effectively identical for DMSO-treated, STM2457-treated and IVT datasets below 0.98,
139 while the profiles differ substantially at probabilities ≥ 0.98. This suggests that sites detected below 0.98
140 are likely false positives, indicating that only high-probability m^6A calls can be considered reliable. Here,

141 STM2457 treatment shows a marked reduction in the number of reported m^6A sites in both all-context
142 and DRACH-only datasets when compared to the DMSO treatment (**Fig. 2A,B**).

143 In addition to m^6A, we examined modification probability distributions for Ψ, inosine and m^5C, each of
144 which can be basecalled with Dorado v0.9.0 (**Fig. S2**). For Ψ and I, the modification probability distributions
145 for the STM2457 and DMSO datasets only diverge from the IVT dataset at ≥ 0.99 modification probability,
146 thus providing a relevant threshold for filtering out false positives (**Fig. S2A,B**). By contrast, m^5C
147 modification probability distributions were similar between all three datasets and no filter could be
148 adequately established (**Fig. S2C**). While this likely reflects a low rate of m^5C installation, it also indicates
149 that the m^5C model utilized by Dorado requires further refinement.

150 Modification probability distributions have similar profiles for transcriptome-aligned NHDF data (**Fig. S3)**
151 as well as genome- and transcriptome-aligned data from HD10.6 cells **(Fig. S4, Fig. S5**). Of note, the fraction
152 of high-probability inosine sites is increased in STM2457-treated cells, compared to DMSO-treated ones
153 in genome-aligned data (**Fig. S2B, Fig.S4D**), which points to increased numbers of retained introns in the
154 METTL3-inhibited sample. Accordingly, this effect is absent in transcriptome-aligned data (**Fig. S3D, Fig.**
155 **S5D**).

156 While IVT transcriptomes serve as excellent negative controls, we also considered whether the RNA
157 calibration strand (RCS) that is added during DRS library preparation could serve a similar function. The
158 RCS is a synthetic Enolase II RNA derived by IVT[13] that, similarly to the IVT transcriptomes, are devoid of
159 RNA modifications. Despite the reduced diversity of sequence in the RCS data, the overall results were
160 comparable to that from the IVT transcriptome data (**Fig. S6A,B, Fig. 2A,B**), demonstrating that the RCS
161 can be used as a "built-in" negative control for modification detection with DRS. Finally, we determined
162 that basecall and modification probability distributions remain generally consistent between Dorado
163 versions although more m^6A and Ψ sites are detected with high modification probability in Dorado versions
164 v0.8.0/v0.9.0 compared to v0.7.0 (**Fig. S6C-E)**.

165 **Filtering by stoichiometry further eliminates potential false positives**

166 We next examined the stoichiometry distributions for all m^6A sites in both unfiltered and filtered NHDF
167 datasets (**Fig. S7**). Here, unfiltered datasets showed that the vast majority of m^6A sites, 92% for all-context
168 and 75% for DRACH-context, have stoichiometry values < 10% with less than 1% sites showing high (> 50%)
169 stoichiometry. Application of our filtering strategy removed over 70% of the sites (**Fig. 2E)**, regardless of
170 the context, almost all of which showed stoichiometries of < 10%, without significantly changing the

171 numbers of high stoichiometry sites detected (**Fig. S7**). A similar analysis of the IVT transcriptome revealed
172 large numbers of false positive m^6A sites with stoichiometry values below 10%, most of which were
173 eliminated by application of our filters (**Fig. S8A,B**). However, the remaining presence of low stoichiometry
174 false-positive sites indicates that an additional layer of filtering to exclude sites with < 10% stoichiometry
175 is beneficial. An additional argument in support of our modification probability filtering strategy is the
176 observation that treatment with the METTL3 specific inhibitor STM2457 dramatically reduced m^6A
177 installation at sites with ≥ 10% stoichiometry in untreated datasets in a context-independent manner (**Fig.
178 2C,D, Supplementary Table 2**). Similar results were observed for the transcriptome-aligned NHDF and
179 HD10.6 m^6A stoichiometry distributions as well (**Fig. S9, Fig. S10, Supplementary Table 2**).

180 To determine whether our filtering strategy could be applied to other RNA modifications detected by
181 Dorado, we generated stoichiometry distributions for Ψ and inosine in unfiltered and filtered (≥ 0.99
182 modification probability) human genome-aligned DMSO- and STM2457-treated NHDFs and HD10.6s. As
183 with m^6A (**Fig. S7**), 92% of detected Ψ and 96% of inosine have less than 10% stoichiometry (**Fig. S11, Fig.
184 S12)**. A recent study that explores Ψ in human mRNAs using 2-bromoacrylamide-assisted cyclization
185 sequencing (BACS) corroborates our results, showing that 82% of Ψ modifications are detected at low
186 levels (below 20%)[29].

187 Above 10% stoichiometry, in our data, there are more inosine sites in STM2457-treated cells than DMSO-
188 treated cells, reflected also in the stoichiometry distributions (**Fig. S11, Fig. S12, Supplementary Table 2**).
189 Ψ and inosine stoichiometry distributions in our IVT data **(Fig.S8C,D)** further validates our filtering strategy
190 of ≥ 0.99 modification probability, as it substantially reduces the number of detected false-positive sites in
191 the unfiltered data even at higher (> 50%) stoichiometries in case of inosine (**Fig.S8D**).

192 We have thus demonstrated that setting thresholds for basecall modification probabilities provides a
193 robust method for identifying and filtering false-positive modification calls. For the purposes of this study,
194 we further evaluated the effect of placing the modification threshold on the modification data and
195 restricted the downstream analyses to m^6A sites with probability scores ≥ 0.98.

196 **RNA modification detection rates differ between Dorado versions**

197 Dorado basecalling relies on models generated using Remora (github.com/nanoporetech/remora) that are
198 frequently updated. This raises an important question as to whether these updates produce notably
199 different outputs that could impact biological interpretation. To address this, we utilized multiple Dorado
200 versions and Remora models to perform modification calling of our NHDF dataset. We restricted our

201 analyses to m^6A and Ψ as these are the only modifications that could be detected in previous versions.
202 Here, we observe that m^6A and Ψ stoichiometry distributions are generally consistent between different
203 Dorado versions (**Fig. S13**), albeit with greater numbers of low stoichiometry (< 10%) sites detected in
204 older Dorado versions. There is a large overlap of the detected ≥ 10%-stoichiometry m^6A (DRACH and all-
205 context) and Ψ sites between versions, although there are a lot of unique all-context m^6A and Ψ sites
206 identified with just v0.7.0 and not v0.8.0-v0.9.0, that are reduced after applying our modification
207 probability filter (**Fig. S14A,B**). Ψ sites uniquely detected with different versions are found in slightly
208 different sequence contexts (**Fig. S14C,D**). Thus, while RNA modification detection with Dorado is sensitive
209 to changes in the underlying models/versions, this can be effectively controlled through the application of
210 filters that only retain sites with high-confidence modification probability calls.

211 **GLORI exhibits reduced specificity but converges with Dorado predictions at high stoichiometry sites**

212 To assess the agreement of DRS Dorado-derived predictions with an orthogonal m^6A detection method,
213 we conducted GLORI on the same NHDF RNA preparations. To ensure a fair comparison of m^6A sites across
214 the two methods, we applied the following filtering strategies: For Dorado (DRACH model), this included
215 thresholds on modification probability (A ≥ 0.8, m^6A ≥ 0.98) and read coverage (≥ 20). For GLORI, we used
216 filters based on coverage (≥ 20) and statistical significance (p-value ≤ 0.05). Next, we extracted m^6A sites
217 that were present in both DMSO and STM2457 treated samples and removed all sites with ≤ 10%
218 stoichiometry in DMSO (**Fig. 3A**). This stepwise strategy reduced false positives and enabled a fair
219 comparison of the high confidence m^6A sites identified by both approaches.

220 As shown in the violin plots, both the all-context and DRACH-context Dorado models identify similar
221 numbers of m^6A sites (81,012 (all-context model) and 72,372 (DRACH model)), the stoichiometries of
222 which were strongly reduced upon STM2457 treatment (**Fig. 3B**). When we analyzed the motifs of the m^6A
223 sites detected with the all-context model, we observed nearly 9,000 sites outside of DRACH motifs.
224 Importantly, the total number of m^6A sites within DRACH motifs was nearly identical to the number of m^6A
225 sites reported using the DRACH model (**Fig.3B**). Unexpectedly, GLORI detects a substantially higher
226 number of sites when using all motifs (140,934) compared to DRACH motifs alone (27,926) (**Fig. 3C**). The
227 fact that most GLORI-identified sites fall outside DRACH motifs suggests these are false positives. This
228 interpretation is reinforced by the observation that the overlap between Dorado and GLORI datasets
229 remains virtually unchanged when comparing GLORI "all motif" and "DRACH motifs" (n=13,614) site sets,
230 underscoring that the additional non-DRACH sites in GLORI (n= 113,008) are largely spurious (**Fig. S15A**).

231 We obtained similar results when comparing m^6A sites obtained with Dorado and GLORI from HD10.6
232 cells. Overall, we detected a similar number of m^6A sites in HD10.6 cells using the all-context and DRACH-
233 specific Dorado models (77,728 and 67,222, respectively) (**Fig. S15B**). By contrast, we detected 174,282 vs
234 27,965 m^6A sites using GLORI without or with DRACH filtering, respectively (**Fig. S15C**).

235 To evaluate how stoichiometry thresholds affect the agreement between GLORI and Dorado, we compared
236 the overlap of high-confidence GLORI-identified sites with Dorado m^6A (DRACH context) calls across
237 increasing m^6A stoichiometry cutoffs (**Fig. 3D**). At the lowest stoichiometry threshold applied (≥ 10%),
238 13,614 sites overlapped, representing less than half (48.8%) of GLORI sites. However, as the stringency
239 increased, the relative overlap increased sharply: at ≥ 20%, nearly 70% of GLORI sites overlapped (11,536
240 sites), at ≥ 50% almost 90% overlapped (6,307 sites), and at the highest cutoff (≥ 80%), over 93% of GLORI
241 sites (3,119 sites) were also detected by Dorado. These results demonstrate that while GLORI detects a
242 broad set of potential m^6A sites, only the sites with high stoichiometries overlap with Dorado, indicating
243 high signal to noise ratios in the GLORI datasets.

244 GLORI relies on the high-efficiency conversion of adenosine to inosine with overall gene A to I conversion
245 rates needing to exceed 95-98%[29,30]. Similarly, we observed the conversion rates for our NHDF and HD10.6
246 datasets to be 95% (Supplementary **Table 3**), but we found that small variations in gene conversion rates
247 can strongly influence the sensitivity and specificity of GLORI, as only high-stoichiometry m^6A sites
248 overlapped with DRS-derived predictions (**Fig. 3D**).

249 **DRS coupled with Dorado outperforms GLORI in isoform-specific m^6A detection**

250 Next, we wanted to highlight the advantage of long-read sequencing in terms of mapping reads with
251 isoform-precision to the human transcriptome, which is not possible using GLORI due to its reliance on
252 short-read sequencing. We first confirmed that mapping Dorado-basecalled sites to the human
253 transcriptome results in the detection of similar numbers of m^6A sites in both NHDFs and HD10.6 cells (**Fig.**
254 **S15D,E**). To assess if DRS coupled with Dorado identifies distinct m^6A profiles on different transcript
255 isoforms, we analyzed the *SPEN* gene locus which was previously shown to contain more than 100 m^6A
256 sites[16]. On a genomic level, we detected 42 m^6A sites using DRS, while GLORI only detected 5 m^6A sites
257 (**Fig. 3E**). Out of the 7 total isoforms of *SPEN*, *SPEN-201* and *SPEN-202* were detectable in DRS-derived
258 data, considering a minimal coverage of 20 (**Supplementary Table 4**). When plotting the m^6A sites detected
259 by transcriptome-aligned DRS data, we found that the *SPEN-201* isoform exhibited more m^6A sites, albeit

260 with a lower median stoichiometry compared to the *SPEN-202* isoform (**Fig. 3F**). This example highlights
261 that DRS is superior compared with GLORI for isoform-level m^6A quantification.

262 **Dorado profiling reveals cell type differences in m^6A distribution and stoichiometry**

263 To assess the reproducibility of Dorado across DRS datasets derived from different cell types, we compared
264 the m^6A methylomes of NHDFs and HD10.6 cells. To ensure a fair comparison, we extracted genes that
265 were expressed in both cell types. Next, we compared the total numbers of high confidence m^6A sites in
266 NHDFs and HD10.6 cells within the genes expressed in both cell types and found that 87% of HD10.6-
267 derived m^6A sites overlap with those detected in NHDFs (**Fig. 4A**). Site-specific m^6A stoichiometries from
268 the shared gene set correlated strongly across cell types (Pearson r = 0.93), although we observed a global
269 shift in the HD10.6 dataset, where m^6A stoichiometries were generally higher than in NHDFs (**Fig. 4B**).
270 Gene Ontology (GO) term analysis of genes containing shared or unique m^6A sites across NHDFs and
271 HD10.6 cells further revealed that these modifications are enriched in transcripts involved in fundamental
272 cellular processes such as transcriptional regulation, DNA repair, chromatin remodeling, and protein
273 transport suggesting that conserved sites maintain core cellular functions (**Fig. 4C**).

274 We next aimed to conduct a comparative motif analysis of m^6A sites detected in NHDFs and HD10.6 cells.
275 Sequence logos confirmed strong enrichment of the canonical DRACH consensus motif for both cell types
276 in modification probability-filtered data, with the GGACU pentamer being the most prominent (**Fig. 4D,E,**
277 **Fig.S16**). Quantification of 5-mer distributions revealed highly similar motif preferences across datasets.
278 While the global motif architecture of m^6A deposition is conserved between NHDFs and HD10.6 cells, we
279 observed a trend toward higher stoichiometries in the HD10.6 dataset, particularly within the most
280 abundant DRACH motifs, suggesting enhanced modification efficiency in the neuronal-like context (**Fig. 4F,**
281 **G**).

282 Metagene analysis showed, as known previously, that m^6A is enriched near the stop codons and in the 3'
283 UTRs (**Fig. S17A**). Since there have been several studies done on how exon architecture regulates m^6A
284 distribution on mRNAs[1,2,3], we also examined m^6A sites in monoexonic transcripts separately and observed
285 that monoexonic transcripts are more equally m^6A-modified across entire CDS (**Fig. S17A**). m^6A (both all-
286 context and DRACH-context) distribution profiles look less typical for all-stoichiometry or ≤ 10%
287 stoichiometry sites in both NHDF and HD10.6 datasets (**Fig. S17B+C**). For all-context m^6A (**Fig. S17B**),
288 modification filter also proves to be deciding for the distribution profile, together with ≥ 10%
289 stoichiometry.

290 To connect shared m^{6}A sites with their functional context and illustrate their distribution on individual
291 transcripts, we conducted a KEGG pathway analysis with the genes that contain shared m^{6}A sites between
292 NHDFs and HD10.6 cells. We observed a significant enrichment in pathways related to neurodegeneration,
293 infection, and intracellular signaling, including axon guidance and mTOR signaling (**Fig. 4H**). To further
294 illustrate these findings, we examined representative genes contained in one or more of these pathways.
295 *WNTB5*, *MAPK1*, *AKT1*, *NCK1* and *NRP1* all showed robust methylation in both NHDFs and HD10.6 cells,
296 with consistently higher stoichiometries observed in the neuronal-like context. In contrast, *ACTG1*,
297 included as a housekeeping control not represented in the enriched pathways, displayed stable and
298 comparatively uniform methylation across both cell types (**Fig. 4I**).

299 Together, these results demonstrate that while the global architecture of m^{6}A deposition is largely
300 conserved between primary fibroblasts and neuronal-like cells, the neuronal context is characterized by a
301 global shift toward higher m^{6}A stoichiometries, which is readily detectable in single genes. These findings
302 suggest that cell type-specific regulation of m^{6}A may fine-tune core cellular pathways relevant to that cell
303 type.

304

305 ## Discussion

306 High rates of false positives remain a major problem in epitranscriptomic mapping tools, even with the
307 advent of modification-aware basecalling models[23]. Comparing the outputs of orthogonal methods
308 performed on the same input material, combined with use of IVT RNA and/or specific methyltransferase
309 inhibition, offers an effective mitigation strategy[17,23]. The results presented here establish a robust
310 framework that minimizes false positives and enables isoform-specific and quantitative m^{6}A mapping.
311 Through a direct comparison with GLORI, we showed that combining DRS datasets with Dorado
312 modification-aware basecalling provides greater scalability, isoform resolution, and reproducibility across
313 different cell types. Further, the simplicity of DRS library preparation enhances consistency whereas GLORI
314 is constrained by a laborious and technically challenging protocol that is extremely sensitive to adenosine
315 conversion rates[31]. Despite these limitations, we determined that the majority of m^{6}A sites in our filtered
316 DRS datasets could be confirmed by GLORI. By expanding our analysis across multiple human cell types,
317 we further determine that the sites of m^{6}A installation are conserved for the most part but that the
318 stoichiometries show cell-type specific differences.

319 Crucially, the inclusion of transcriptome-wide IVT datasets and the use of a selective METTL3 inhibitor
320 (STM2457) provides a robust approach to developing filters that dramatically decrease false-positive m^6A
321 detection. The impact of both approaches was readily visualized by plotting modification probability
322 distributions, enabling our determination that retaining only m^6A calls with confidence scores of 0.98 or
323 greater retained almost all STM2457-sensitive m^6A sites while eliminating the large numbers of false-
324 positive m^6A sites reported for the IVT datasets. This approach is supported by a recent study which
325 highlighted that Dorado introduces large numbers of false positives outside canonical DRACH contexts[32].
326 Our data confirms that in the unfiltered datasets, the number of false positives is higher in the all-context
327 model compared to the DRACH model. This may reflect the degenerate nature of the canonical DRACH
328 consensus and understudied differences in sequence preference between different cell types and
329 organisms. Retaining only sites at which m^6A stoichiometries exceeded 10% further reduced false-positive
330 rates while also focusing the results on identification of sites that are more likely to be biologically
331 significant. The combined effect of these filters was to provide a strong agreement between the m^6A all-
332 context and m^6A DRACH model in the detection of high confidence m^6A sites, as observed by filtering for
333 DRACH motifs from m^6A sites detected with the all-context model (**Fig.3B**).

334 Prior studies have emphasized the need for isoform-resolved m^6A quantification by demonstrating
335 widespread heterogeneity in m^6A levels between genes and their isoforms[32,33]. Here, we showed that
336 aligning DRS datasets to the human transcriptome reveals differential m^6A installation patterns between
337 isoforms, underscoring the importance of accurate, isoform-specific m^6A quantification at single-molecule
338 resolution and highlighting the advantage of DRS against short-read sequencing approaches such as GLORI
339 (**Fig. 3E,F**).

340 Recently, the ability of Dorado and m6Anet[34] to identify m^6A in DRS datasets from 293T and HeLa cells
341 were systematically evaluated using IVT RNA and comparisons to published datasets obtained by GLORI
342 and eTAM-seq[22]. In this case the authors found that Dorado (v0.7.2) outperforms m6Anet but also
343 advocated for the importance of applying filters to reduce false positive calls. Likewise, we found that IVT
344 RNA served as a valuable negative control to detect false positive calls. Furthermore, using the RNA
345 calibration strand (RCS) as a negative control provided a convenient and simpler alternative to IVT RNA,
346 yielding comparable probability distributions and thereby enabling a built-in control for routine
347 experiments[20]. While m^6A detection was generally reproducible across Dorado versions, model-specific
348 differences are present, i.e. v0.7.0 produces substantially more low-confidence all-context m^6A and Ψ calls

349 than v0.8.0–0.9.0, highlighting that the models remain under active development and that benchmarking
350 should be repeated when switching to a new model.

351 Focusing on other modifications that are detected by Dorado, we observe that inosine and Ψ detection
352 appears robust when applying similar filtering strategies as for m^6A, the latter supported by other
353 studies[24,35]. For inosine, we observed an enrichment of high-probability (> 0.99) inosine sites in genome-
354 aligned data from STM2457-treated samples. Further studies are needed to determine whether this
355 reflects increased intron retention or misclassification of modifications by the current inosine model.
356 While DRS is the only approach that allows for robust isoform-specific m^6A mapping, comparisons with
357 orthogonal approaches remain essential to validate findings. When comparing Dorado and GLORI using
358 the same input RNA, we unexpectedly observed that only 50% of m^6A sites with stoichiometries above
359 10% that were identified by Dorado were also present in the GLORI data although increasing the
360 stoichiometries led to increased overlap (**Fig.3D**). We attribute this to a sub-optimal gene-wide adenosine
361 conversion rate of 95% during GLORI, compared to the 98-99% rate reported in the original study[36]. Here,
362 unconverted adenosines result in high signal-to-noise ratios which result in lower sensitivity, explaining
363 why the overlap with Dorado increases at higher-stoichiometry sites. This is also reflected in the relatively
364 high amount of false positive sites detected outside of DRACH motifs in the GLORI datasets (**Fig. 3C**). The
365 sensitivity of GLORI to small fluctuations thus limits its utility in comparison to DRS. Whether this can be
366 resolved with improvements to the GLORI methodology[37] requires further study.

367 Prior analyses of m^6A stoichiometry distributions across immortalized human cell lines, including cancer
368 cell lines, have revealed broadly similar m^6A stoichiometry distributions[33]. By contrast, our analysis of
369 primary fibroblasts (NHDFs) and neuronal cells (differentiated HD10.6 cells) revealed an overall increase
370 in m^6A stoichiometries in neurons compared to fibroblasts, especially in genes associated with neuronal
371 function. While further studies are warranted to establish the underlying mechanisms and functional
372 consequences of these differences, this aligns with the importance of m^6A in neuronal function[38]. This
373 includes m^6A maps from synaptic and nuclear RNA isolated from mouse dorsal hippocampus which
374 showed that synapse-specific RNAs showed generally higher methylation levels[37]. Intriguingly, neuronal
375 regulation pathways utilize differentially spliced isoforms[39] and a recent study revealed isoform-specific
376 differences in m^6A sites and stoichiometries in the human brain that are linked to m^6A machinery
377 expression and cell type context[40].

378 In summary, using matched RNA samples from different experimental conditions and from different
379 human cell types analyzed with both Dorado and GLORI orthogonal methodologies, we establish a robust

380 pipeline for the precise and quantitative transcriptome-wide detection of m^6A in DRS datasets. Future
381 studies should aim at extending this approach using orthogonal sequencing approaches to benchmark
382 Dorado-based detection of Ψ, m^5C and inosine, as well as the newly released models for 2'O-methylation.

383

384 # Methods

385 **Cell culturing and STM2457 treatment**

386 NHDFs (Lonza) were cultured in Dulbecco's Modified Eagle Medium (DMEM, Gibco), supplemented with
387 5% heat-inactivated FBS (HI-FBS, Gibco) and 1% penicillin-streptomycin (PS, Lonza) under standard culture
388 conditions. The HD10.6 cell line was originally derived from fetal human dorsal root ganglia (DRG) and
389 immortalized by overexpression of the v-myc oncogene under tetracycline regulation. Upon addition of
390 tetracycline and neurotrophins, these cells can be differentiated into neuron-like cells that show typical
391 neuronal morphology and nociceptive properties[25]. For our experiments, we used HD10.6 cells that have
392 been matured for 10 days. For maintenance of undifferentiated HD10.6 cells (a gift from Anna Cliffe,
393 University of Virginia School of Medicine), T-75 flasks were coated with 17 µg/mL fibronectin in PBS for 15
394 min at 37°C before adding the cell suspension. Cells were grown in Advanced DMEM/F12 (Gibco),
395 supplemented with 1x Neurocult SM1 (STEMCELL Technologies), 2 mM L-glutamine (Gibco), 1x Primocin
396 (Invivogen) and 10 ng/mL prostaglandin E1 (Sigma). 0.5 ng/mL bFGF (PeproTech) were added freshly to
397 each media change. For maturation, HD10.6 cells were seeded at a density of 25.000 cells/cm² in 6-well
398 plates which were pre-coated overnight with 50 µg/mL poly-L-ornithine hydrobromide (Sigma) in 0.5M
399 borate buffer pH 8.5 (Boston Bioproducts), and 1 µg/mL fibronectin (overnight). After cell attachment
400 overnight, media was changed to Neurobasal media (Gibco), supplemented with 1x Neurocult SM1, 2mM
401 L-glutamine, 1x Primocin, 1 µg/mL doxycycline, 50 ng/mL NGF (Alomone Labs), and 25 ng/mL CNTF, GDNF
402 and NT-3 (PeproTech). Media was changed every three days until HD10.6 cells showed complete neuronal
403 morphology after 12 days. NHDFs and matured HD10.6 cells were treated with 30 µM STM2457 (Sigma)
404 or DMSO for 48h, harvested in TRIzol (Invitrogen), and stored at -80°C until RNA isolation.

405 **RNA isolation**

406 RNA was extracted by the TRIzol method according to the manufacturer's instructions. Briefly, 0.2 mL
407 chloroform per 1 mL TRIzol were added and mixed by vortexing. After centrifugation for 30 min at 12,000xg
408 at 4°C, the aqueous phase was transferred to isopropanol (0.5 mL per 1 mL TRIzol), mixed and incubated

409 for 10 min at room temperature. Subsequently, the samples were centrifuged for 10 min at 12,000xg at
410 4°C and the RNA pellets were washed twice in 75% fresh ethanol. After air-drying for 5-10 min, RNA was
411 dissolved in nuclease-free H_2O and stored at -80°C until further processing. One third of each sample was
412 used for DRS and the remaining same-input material was processed for GLORI. For both DRS and GLORI,
413 poly(A) RNA was selected using the Dynabeads™ mRNA Purification Kit (Invitrogen), according to the
414 manufacturer's instructions.

415 **GLORI-Sequencing**

416 GLORI reactions were largely performed following the published protocol for RNA protection, deamination
417 and deprotection with several minor adjustments[31]. Briefly, poly(A) RNA was fragmented at 94°C for 2 min
418 (NEB E6150S) and subjected to glyoxal protection, sodium-nitrite-mediated deamination (750 mM $NaNO_2$
419 at 16°C for 8h and 4°C for 6h), and deprotection in triethylammonium acetate-formamide buffer at 95°C
420 for 10 min. Library construction was conducted using a ligation-based strategy as described by Liu and
421 colleagues incorporating 5'adapters with 11-nt UMIs[41].

422 Indexed libraries were amplified with NEBNext Dual Primers Set 2 (E7780S), purified using AMPure XP
423 beads, and validated on an Agilent TapeStation (D1000 High Sensitivity). Sequencing was performed on an
424 Illumina NovaSeq 6000 platform (150 bp paired-end mode; > 80 million reads per sample).

425 For data processing, we employed the same computational pipeline described in Liu et al[41], but
426 implemented within a Snakemake-based framework that automates the complete analysis starting from
427 raw fastq files[42]. Details on software versions, workflow availability, and repository links are provided in
428 the Data and code availability section below.

429 **Generation of a transcriptome-wide IVT RNA library**

430 We sequenced the IVT RNA from both NHDFs and HD10.6 cells along with the native poly(A) RNA to
431 evaluate potential false-positive predictions by Dorado. This was achieved by converting the poly(A) RNA
432 to cDNA using a poly(A) primer and a T7-incorporating strand switch primer and using the cDNA as a
433 template for IVT synthesis.

434 **Nanopore Sequencing of poly(A) RNA**

435 250-300 ng of isolated poly(A) RNA and 25 ng RNA calibration strand (RCS) were used as input for the
436 standard DRS SQK-RNA004 protocol. Resulting libraries were loaded onto individual PromethION RNA flow
437 cells before sequencing on the PromethION 2 Solo instrument for 48h.

Nanopore basecalling

Basecalling was performed using Dorado v0.5.0, v0.6.0, v0.7.0, v0.8.0 and v0.9.0 (ONT) in super accuracy mode with modification detection enabled for m^6A (DRACH context) in v.0.5.0, 0.6.0, 0.8.0 and 0.9.0, m^6A (all context) in v0.7.0, 0.8.0 and 0.9.0, Ψ in v0.7.0, v0.8.0 and v0.9.0, and inosine and m^5C in v.0.9.0. Each dataset was processed with adapter trimming. Resulting unaligned BAM files were converted to FASTQ files containing modification tags for every read in the FASTQ read headers, using SAMtools v1.18[43] [*samtools fastq -TMM,ML*].

Genome and transcriptome level alignments

For genome-level alignment, reads were aligned against a combined human genome reference (GRCh38.p14), and the RCS gene sequence, using minimap2 v2.26 [*minimap2 -ax splice -k14 -y -uf --secondary=no*]. For transcriptome-level alignment, the reads were aligned against the human transcriptome (v47), obtained via GENCODE[44] [*minimap2 -t 8 -ax map-ont -y -L -p 0.99*]. Using SAMtools v1.18, resulting genome- and transcriptome-aligned SAM files were converted to BAM files retaining only primary alignments [*samtools view -b -F2308* and *samtools view -b -F2324,* respectively]. BAM files were sorted [*samtools sort*] and indexed [*samtools index*].

Processing of RNA modification calls using Modkit

Probability distributions of the detected sites in human genome- and transcriptome-aligned reads were generated using sample-probs subcommand from Modkit v0.4.1[28] [sample-probs --hist --only-mapped --percentiles 0.1,0.2,0.3,0.4,0.5,0.6,0.7,0.8,0.85,0.9 -f 0.1]. For RCS reads -f value was increased to 0.25. The resulting .tsv files were imported into RStudio (v. 2024.12.1+563) and plotted using a custom R(v. 4.4.2) script (Dorado_mod_probability_distributions.R), with the following packages: dplyr (v1.1.4), ggplot2 (v3.5.1), reshape2 (v1.4.4), patchwork (v1.3.0).

To generate extended bedMethyl files with m^6A-modified and unmodified base counts from the reads mapping against the potential modification sites, we used Modkit pileup subcommand with the thresholds [*--filter-threshold A:0.8 --filter-threshold T:0.8 --filter-threshold C:0.8 --filter-threshold G:0.8 --motif A 0 --ref $REF*] with and without modification threshold filter [*--mod-thresholds a:0.98*] to generate our "filtered" and "unfiltered" datasets, respectively. Canonical nucleotide calls (A, C, G, U) were retained with > 80% probability scores for all modification datasets. Modification thresholds were set at 98% for m^6A and 99% for all the other modifications. *--motif* was adjusted to the base for which modification status was examined, which further restricted the analysis to only the positions that matched canonical

468 nucleotides in the reference genome. The resulting sites were further filtered according to the read
469 number per site (≥ 20 reads) for all the downstream analysis.

470 **Gene Ontology analysis**

471 GO analysis was performed using DAVID[45]. Genes harboring m^6A sites were filtered for genes that were
472 generally present in both datasets. Next, they were divided into NHDF-specific, HD10.6-specific, and
473 shared categories, and enrichment was assessed for biological processes. GO terms were considered
474 significant at FDR ≤ 0.05 with a minimum of 20 genes per GO term. For plotting, the top 20 GO terms that
475 contained the highest gene numbers were extracted from the analysis of the shared gene sets.

476 **Analysis and visualization of modification stoichiometries and distributions**

477 Stoichiometry distribution histograms and metaplots were generated using Rstudio and R v4.4.2 with the
478 following libraries: data.table, patchwork, scales, tidyr, dplyr, ggplot2[4]. Access to scripts used for analysis
479 and visualization is detailed below.

480 **Data and code availability**

481 Raw fastq and pod5 datasets generated in this study are available via the European Nucleotide Archive
482 (ENA) under the accession number PRJEB101380.
483 R scripts used in the analysis and visualization of data presented herein are available in the
484 https://github.com/DepledgeLab/HOLDEN repository.
485 The Snakemake workflow implementing the GLORI analysis pipeline is available in the
486 https://github.com/gp-micro/glori repository.

487

488 **Acknowledgements and funding**

489 DPD is supported by a German Centre for Infection Research (DZIF) Associate Professorship and the NIAID
490 grants R01-AI170583 and R01-AI152543. DPD also receives funding from the Deutsche
491 Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy - EXC
492 2155, project number 390874280. EL is supported by the NIAID grant R01-AI170583 as well as the
493 Hannover Biomedical Research School (HBRS) and the Center for Infection Biology (ZIB). DO is supported
494 by the Walter Benjamin Programme of the DFG, project 554758329. ACW is supported by NIAID grants
495 R01AI176335 and R01-AI170583.

References

1 Gilbert, W. V. & Nachtergaele, S. mRNA Regulation by RNA Modifications. *Annu Rev Biochem* **92**, 175-198 (2023). https://doi.org:10.1146/annurev-biochem-052521-035949

2 Cappannini, A. *et al.* MODOMICS: a database of RNA modifications and related information. 2023 update. *Nucleic Acids Research* **52**, D239-D244 (2023). https://doi.org:10.1093/nar/gkad1083

3 Anreiter, I., Mir, Q., Simpson, J. T., Janga, S. C. & Soller, M. New Twists in Detecting mRNA Modification Dynamics. *Trends Biotechnol* **39**, 72-89 (2021). https://doi.org:10.1016/j.tibtech.2020.06.002

4 Yang, L. *et al.* Emerging role of RNA modification and long noncoding RNA interaction in cancer. *Cancer Gene Ther* **31**, 816-830 (2024). https://doi.org:10.1038/s41417-024-00734-2

5 Zaccara, S., Ries, R. J. & Jaffrey, S. R. Reading, writing and erasing mRNA methylation. *Nature Reviews Molecular Cell Biology* **20**, 608-624 (2019). https://doi.org:10.1038/s41580-019-0168-5

6 Murakami, S. & Jaffrey, S. R. Hidden codes in mRNA: Control of gene expression by m(6)A. *Mol Cell* **82**, 2236-2251 (2022). https://doi.org:10.1016/j.molcel.2022.05.029

7 He, P. C. & He, C. m(6) A RNA methylation: from mechanisms to therapeutic potential. *Embo j* **40**, e105977 (2021). https://doi.org:10.15252/embj.2020105977

8 Sendinc, E. & Shi, Y. RNA m6A methylation across the transcriptome. *Mol Cell* **83**, 428-441 (2023). https://doi.org:10.1016/j.molcel.2023.01.006

9 Sun, H., Li, K., Liu, C. & Yi, C. Regulation and functions of non-m(6)A mRNA modifications. *Nat Rev Mol Cell Biol* **24**, 714-731 (2023). https://doi.org:10.1038/s41580-023-00622-x

10 Moshitch-Moshkovitz, S. *et al.* mRNA m6A detection. *Nature Reviews Methods Primers* **4** (2024). https://doi.org:10.1038/s43586-024-00365-9

11 Dominissini, D. *et al.* Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature* **485**, 201-206 (2012). https://doi.org:10.1038/nature11112

12 Meyer, Kate D. *et al.* Comprehensive Analysis of mRNA Methylation Reveals Enrichment in 3' UTRs and near Stop Codons. *Cell* **149**, 1635-1646 (2012). https://doi.org:https://doi.org/10.1016/j.cell.2012.05.003

13 Linder, B. *et al.* Single-nucleotide-resolution mapping of m6A and m6Am throughout the transcriptome. *Nature Methods* **12**, 767-772 (2015). https://doi.org:10.1038/nmeth.3453

14 Meyer, K. D. DART-seq: an antibody-free method for global m6A detection. *Nature Methods* **16**, 1275-1280 (2019). https://doi.org:10.1038/s41592-019-0570-0

15 Xiao, Y. L. *et al.* Transcriptome-wide profiling and quantification of N(6)-methyladenosine by enzyme-assisted adenosine deamination. *Nat Biotechnol* **41**, 993-1003 (2023). https://doi.org:10.1038/s41587-022-01587-6

16 Liu, C. *et al.* Absolute quantification of single-base m(6)A methylation in the mammalian transcriptome using GLORI. *Nat Biotechnol* **41**, 355-366 (2023). https://doi.org:10.1038/s41587-022-01487-9

17 Diensthuber, G. & Novoa, E. M. Charting the epitranscriptomic landscape across RNA biotypes using native RNA nanopore sequencing. *Mol Cell* **85**, 276-289 (2025). https://doi.org:10.1016/j.molcel.2024.12.014

18 Garalde, D. R. *et al.* Highly parallel direct RNA sequencing on an array of nanopores. *Nat Methods* **15**, 201-206 (2018). https://doi.org:10.1038/nmeth.4577

19 Jain, M., Olsen, H. E., Paten, B. & Akeson, M. The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biology* **17**, 239 (2016). https://doi.org:10.1186/s13059-016-1103-0

20 Viehweger, A. *et al.* Direct RNA nanopore sequencing of full-length coronavirus genomes provides novel insights into structural variants and enables modification analysis. *Genome Res* **29**, 1545-1554 (2019). https://doi.org:10.1101/gr.247064.118

21 Abebe, J. S., Verstraten, R. & Depledge, D. P. Nanopore-Based Detection of Viral RNA Modifications. *mBio* **13**, e03702-03721 (2022). https://doi.org:doi:10.1128/mbio.03702-21

22 Zou, Y. *et al.* A comparative evaluation of computational models for RNA modification detection using nanopore sequencing with RNA004 chemistry. *Briefings in Bioinformatics* **26** (2025). https://doi.org:10.1093/bib/bbaf404

23 Cruciani, S. *et al.* De novo basecalling of RNA modifications at single molecule and nucleotide resolution. *Genome Biology* **26**, 38 (2025). https://doi.org:10.1186/s13059-025-03498-6

24 Esfahani, N. G., Stein, A. J., Akeson, S., Tzadikario, T. & Jain, M. Evaluation of Nanopore direct RNA sequencing updates for modification detection. *bioRxiv* (2025). https://doi.org:10.1101/2025.05.01.651717

25 Raymon, H. K. *et al.* Immortalized Human Dorsal Root Ganglion Cells Differentiate into Neurons with Nociceptive Properties. *The Journal of Neuroscience* **19**, 5420-5428 (1999). https://doi.org:10.1523/jneurosci.19-13-05420.1999

26 Yankova, E. *et al.* Small-molecule inhibition of METTL3 as a strategy against myeloid leukaemia. *Nature* **593**, 597-601 (2021). https://doi.org:10.1038/s41586-021-03536-w

27 Yang, Y. *et al.* Current progress in strategies to profile transcriptomic m(6)A modifications. *Front Cell Dev Biol* **12**, 1392159 (2024). https://doi.org:10.3389/fcell.2024.1392159

28 *https://github.com/nanoporetech/modkit*.

29 Xu, H. *et al.* Absolute quantitative and base-resolution sequencing reveals comprehensive landscape of pseudouridine across the human transcriptome. *Nat Methods* **21**, 2024-2033 (2024). https://doi.org:10.1038/s41592-024-02439-8

30 Xie, Y.-Y. *et al.* Single-molecule direct RNA sequencing reveals the shaping of epitranscriptome across multiple species. *Nature Communications* **16**, 5119 (2025). https://doi.org:10.1038/s41467-025-60447-4

31 Shen, W. *et al.* GLORI for absolute quantification of transcriptome-wide m(6)A at single-base resolution. *Nat Protoc* **19**, 1252-1287 (2024). https://doi.org:10.1038/s41596-023-00937-1

32 Xie, Y. Y. *et al.* Single-molecule direct RNA sequencing reveals the shaping of epitranscriptome across multiple species. *Nat Commun* **16**, 5119 (2025). https://doi.org:10.1038/s41467-025-60447-4

33 Guo, W. *et al.* Single-molecule m6A detection empowered by endogenous labeling unveils complexities across RNA isoforms. *Molecular Cell* (2025). https://doi.org:https://doi.org/10.1016/j.molcel.2025.01.014

34 Hendra, C. *et al.* Detection of m6A from direct RNA sequencing using a multiple instance learning framework. *Nature Methods* **19**, 1590-1598 (2022). https://doi.org:10.1038/s41592-022-01666-1

35 Verstraten, R. *et al.* Defining expansions and perturbations to the RNA polymerase III transcriptome and epitranscriptome by modified direct RNA nanopore sequencing. *bioRxiv* (2025). https://doi.org:10.1101/2025.03.07.641986

36 Shen, W. *et al.* GLORI for absolute quantification of transcriptome-wide m6A at single-base resolution. *Nature Protocols* **19**, 1252-1287 (2024). https://doi.org:10.1038/s41596-023-00937-1

37 Sun, H. *et al.* Mild and ultrafast GLORI enables absolute quantification of m6A methylome from low-input samples. *Nature Methods* **22**, 1226-1236 (2025). https://doi.org:10.1038/s41592-025-02680-9

38 Yu, J., She, Y. & Ji, S.-J. m6A Modification in Mammalian Nervous System Development, Functions, Disorders, and Injuries. *Frontiers in Cell and Developmental Biology* **Volume 9 - 2021** (2021). https://doi.org:10.3389/fcell.2021.679662

39 Su, C.-H., D, D. & Tarn, W.-Y. Alternative Splicing in Neurogenesis and Brain Development. *Frontiers in Molecular Biosciences* **Volume 5 - 2018** (2018). https://doi.org:10.3389/fmolb.2018.00012

40 Gleeson, J. *et al.* Isoform-level profiling of m(6)A epitranscriptomic signatures in human brain. *Sci Adv* **11**, eadp0783 (2025). https://doi.org:10.1126/sciadv.adp0783

41 Liu, J. F., Hawley, B. R., Nicholson, L. S. & Jaffrey, S. R. Decoding m(6)Am by simultaneous transcription-start mapping and methylation quantification. *Elife* **13** (2025). https://doi.org:10.7554/eLife.104139

42 Köster, J. & Rahmann, S. Snakemake--a scalable bioinformatics workflow engine. *Bioinformatics* **28**, 2520-2522 (2012). https://doi.org:10.1093/bioinformatics/bts480

43 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079 (2009). https://doi.org:10.1093/bioinformatics/btp352

44 Mudge, J. M. *et al.* GENCODE 2025: reference gene annotation for human and mouse. *Nucleic Acids Res* **53**, D966-d975 (2025). https://doi.org:10.1093/nar/gkae1078

45 Huang da, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**, 44-57 (2009). https://doi.org:10.1038/nprot.2008.211

Figure legends

Figure 1. Workflow for m^{6}A mapping in different human cell types using orthogonal sequencing approaches.

(A) Schematic of the experimental setup. RNA was isolated from NHDFs and differentiated HD10.6 cells treated with STM2457 or DMSO for 48h. Poly(A) RNA was subjected to DRS and GLORI from the same input material. IVT RNA was generated in parallel to assess potential false-positive modification calls. **(B)** Workflow for benchmarking Dorado basecalling. Reads were processed with Dorado versions v.0.5.0 through v0.9.0, aligned to the human genome or transcriptome, and analyzed with ONT's Modkit tool. Multiple filtering strategies were tested against IVT controls and METTL3 inhibition (STM2457). Resulting m^{6}A sites were compared with GLORI data from the same input RNA. **(C)** Final benchmarked Dorado pipeline to detect high-confidence m^{6}A sites.

623 **Figure 2. Modification probability of 0.98 and stoichiometry score of ≥ 10% are necessary cutoffs to**
624 **identify m^{6}A sites accurately.**

625 **(A)** Dorado v0.9.0-generated modification probability distributions for all-context m^{6}A from genome-
626 aligned DMSO- and STM2457-treated NHDFs plotted against the ones from IVT NHDF poly(A) RNA. DMSO
627 probability score distribution is shown as a red solid line, STM2457 – red dashed line, IVT – grey solid line.
628 **(B)** Same as in (A), but for DRACH-context m^{6}A sites. **(C)** Stoichiometry (≥ 10%) distributions for all-context
629 m^{6}A from DMSO- (in red) or STM2457-treated (in pink) NHDF datasets (genome-aligned) are plotted using
630 a 0.98 modification probability cutoff and a filter for ≥ 20 reads. **(D)** Same as in (C), but for DRACH-context
631 m^{6}A sites. **(E)** Overlap of m^{6}A sites between all-context and DRACH models (A basecall accuracy ≥ 80%,
632 coverage ≥ 20reads, stoichiometry ≥ 10%), comparing non-filtered to filtered (modification probability >
633 0.98) outputs. **(F)** Metagene plots showing the density of m^{6}A (all-context – dark red, DRACH-context –
634 bright red) sites across all annotated poly(A) transcripts (left) and monoexonic poly(A) transcripts (right)
635 from genome-aligned DMSO-treated NHDF datasets filtered with 0.98 modification probability, ≥ 20
636 coverage and ≥ 10% stoichiometry.

637

638 **Figure 3. Comparison of m^{6}A sites detected by Dorado and GLORI.**

639 **(A)** Filtering strategy for Dorado vs GLORI to obtain a high-confidence list of m^{6}A sites in NHDFs. **(B)**
640 Stoichiometry distributions are plotted from genome-aligned DMSO- or STM2457-treated NHDF datasets
641 using the indicated Dorado models and the filtering strategy shown in (A). **(C)** Stoichiometry distributions
642 from GLORI performed on DMSO- or STM2457-treated NHDFs are plotted, using the filtering strategy
643 shown in (A). In addition, m^{6}A sites were filtered for DRACH motifs (right plot). **(D)** Overlap of high-
644 confidence m^{6}A sites detected by Dorado vs GLORI. GLORI datasets were filtered using increasing m^{6}A
645 stoichiometry cutoffs (from left to right). **(E)** m^{6}A sites in the *SPEN* gene detected by GLORI and Dorado.
646 Boxplots (right) represent the m^{6}A stoichiometries per m^{6}A site shown on the left. **(F)** *SPEN* isoforms with
647 m^{6}A sites, determined by Dorado. Boxplots (right) show the distribution of m^{6}A stoichiometries on the
648 m^{6}A sites on *SPEN-201* and *SPEN-202*.

649

650 **Figure 4. Comparative analysis of m^{6}A modification patterns in NHDFs and HD10.6 cells using Dorado**
651 **v0.9.0.**

652 **(A)** Overlap of high-confidence m^6A sites between NHDFs and HD10.6 cells. **(B)** Correlation of shared m^6A
653 stoichiometries (n=38,217). **(C)** GO term analysis (Biological Process, DAVID Bioinformatics Resources) of
654 genes with shared or unique m^6A sites. **(D+E)** Motif analysis showing m^6A (all-context and DRACH-context)
655 motif logos, 5-mer distributions and modification stoichiometries in NHDFs. **(F+G)** Same as in (D+E) but for
656 HD10.6 cells. **(H)** KEGG pathway analysis (DAVID Bioinformatics Resources) of shared m^6A-containing
657 genes. **(I)** m^6A stoichiometries of representative examples of individual genes from enriched pathways (H).
658 *ACTG1* serves as a housekeeping control.
659

660 **Figure S1. Evaluation of Dorado by base and modification probability distributions.**

661 **(A)** Dorado v0.9.0-generated basecall accuracy probability values for A (all-context and DRACH-context
662 basecaller models) from genome-aligned DMSO- and STM2457-treated NHDF samples against IVT poly(A)
663 NHDF RNA. A probability score distribution is shown as a black solid line for DMSO, STM2457 – black
664 dashed line, IVT – grey solid line. **(B)** Dorado v0.9.0-generated modification probability distributions,
665 visualized as histograms, for DRACH-context m^6A from genome-aligned DMSO- and STM2457-treated
666 NHDFs and IVT NHDFs. IVT modification probability score distribution is shown in grey, DMSO – bright red,
667 STM2457 – dark red. **(C)** Same as in (B), but for all-context m^6A sites.

668

669 **Figure S2. Modification probability cutoffs are necessary to identify Ψ and inosine sites more accurately.**

670 **(A)** Dorado v0.9.0-generated basecall accuracy probability values for U (left) and modification probability
671 distributions for Ψ (right) from genome-aligned DMSO- and STM2457-treated NHDF samples are plotted
672 against IVT poly(A) RNA. Ψ probabilities are visualized in blue solid and dashed lines for DMSO and
673 STM2457, respectively and in grey solid line for IVT. **(B)** Same as in (A), but showing A base accuracy
674 probabilities and inosine sites. Inosine probabilities are visualized in purple solid and dashed lines for
675 DMSO and STM2457, respectively and in grey solid line for IVT. **(C)** Same as in (A), but showing C base
676 accuracy probabilities and m^5C sites. m^5C probabilities are visualized in green solid and dashed lines for
677 DMSO and STM2457, respectively and in grey solid line for IVT. **(D+E)** DRACH-context and all-context m^6A
678 modification probability distributions (red) from genome-aligned DMSO- and STM2457-treated NHDFs are
679 plotted against RCS negative control (grey).

680

681 **Figure S3. Modification probability cutoffs are necessary to identify modification sites accurately, shown**
682 **for transcriptome-aligned data.**

683 Dorado v0.9.0-generated basecall accuracy probability values for A, U and C (left, **A-E**) and modification
684 probability distributions for m^6A (all-context), m^6A (DRACH-context), Ψ, inosine and m^5C (right, **A-E**) in
685 transcriptome-aligned DMSO-and STM2457-treated NHDFs, plotted against the values from IVT NHDF
686 poly(A) RNA.

687

688 **Figure S4. Modification probability distribution profiles in genome-aligned DMSO-and STM2457-treated**
689 **HD10.6 cells.**

690 Dorado v0.9.0-generated basecall accuracy probability values for A, U and C (left, **A-E**) and modification
691 probability distributions for m^6A (all-context), m^6A (DRACH-context), Ψ, inosine and m^5C (right, **A-E**) in
692 human genome-aligned DMSO-and STM2457-treated HD10.6 cells, plotted against the values from IVT
693 NHDF poly(A) RNA.

694

695 **Figure S5. Modification probability distribution profiles in transcriptome-aligned DMSO-and STM2457-**
696 **treated HD10.6 cells.**

697 Dorado v0.9.0-generated basecall accuracy probability values for A, U and C (left, **A-E**) and modification
698 probability distributions for m^6A (all-context), m^6A (DRACH-context), Ψ, inosine and m^5C (right, **A-E**) in
699 human transcriptome-aligned DMSO-and STM2457-treated HD10.6 cells, plotted against the values from
700 IVT NHDF poly(A) RNA.

701

702 **Figure S6. Modification site probabilities detected with different versions of Dorado and compared to**
703 **RCS.**

704 **(A+B)** DRACH-context and all-context m^6A modification probability distributions from genome-aligned
705 DMSO (solid red line)- and STM2457 (dashed red line)-treated NHDFs are plotted against RCS negative
706 control (grey). **(C)** Modification probability distributions for all-context m^6A (Dorado v0.7.0, v0.8.0 and
707 v0.9.0), the bottom panel showing the distributions in 0.95-1.00 modification range. **(D)** Same as in (C),
708 but for DRACH-context m^6A (Dorado v0.5.0, v0.6.0, v0.8.0 and v0.9.0). **(E)** Same as in (C), but for Ψ (Dorado

v0.7.0, v0.8.0, v0.9.0). All the plots in C-E are generated from genome-aligned DMSO-treated NHDF datasets and are plotted against RCS negative controls.

Figure S7. Evaluation of Dorado by stoichiometry distributions.

(A) Stoichiometry distributions for all-context m^6A from DMSO- (in red) or STM2457-treated (in pink) NHDF datasets (genome-aligned and filtered for ≥ 20 reads) are plotted, comparing 0.98 modification probability-unfiltered (left) and filtered (right) sites. **(B)** Same as in (A), but for DRACH-context m^6A sites.

Figure S8. Stoichiometry distributions of m^6A, Ψ and inosine in IVT datasets.

(A) All-context m^6A stoichiometry distributions (all stoichiometries (top) and ≥ 10% stoichiometries (bottom)) in (genome-aligned) IVT RNA from DMSO-treated NHDFs, comparing modification probability-unfiltered and filtered data. The bottom panel shows stoichiometry distributions only for ≥ 10% stoichiometry sites. **(B)** As in (A), but for DRACH-context m^6A. **(C)** As in (A), but for Ψ. **(D)** As in (A), but for inosine.

Figure S9. Evaluation of Dorado by stoichiometry distributions in transcriptome-aligned NHDFs.

(A) Stoichiometry distributions for all-context m^6A from DMSO- (in red) or STM2457-treated (in pink) NHDF datasets (transcriptome-aligned and filtered for ≥ 20 reads) are plotted, assessing the 0.98 modification probability cutoff. The bottom panel shows stoichiometry distributions only for ≥ 10% stoichiometry sites. **(B)** Same as in (A), but for DRACH-context m^6A sites.

Figure S10. Evaluation of Dorado by stoichiometry distributions in transcriptome-aligned HD10.6 cells.

(A) Stoichiometry distributions for all-context m^6A from DMSO- (in red) or STM2457-treated (in pink) HD10.6 datasets (transcriptome-aligned and filtered for ≥ 20 reads) are plotted, assessing the 0.98 modification probability cutoff. The bottom panel shows stoichiometry distributions only for ≥ 10% stoichiometry sites. **(B)** Same as in (A), but for DRACH-context m^6A sites.

736 **Figure S11 Stoichiometry distributions of Ψ and inosine in NHDFs.**

737 **(A)** Distributions of all (left) and ≥ 10% (right) Ψ stoichiometries in the 0.99 modification probability-
738 unfiltered and filtered data from genome-aligned DMSO- (darker color) and STM2457-treated (lighter
739 color) NHDFs (filtered for ≥ 20 reads). **(B)** Same as in (A), but for inosine.

740

741 **Figure S12. Stoichiometry distributions of Ψ and inosine in HD10.6 cells.**

742 **(A)** Distributions of all and ≥ 10% Ψ stoichiometries in the 0.99 modification probability-unfiltered and
743 filtered data from genome-aligned DMSO- (darker color) and STM2457-treated (lighter color) HD10.6 cells
744 (filtered for ≥ 20 reads). **(B)** Same as in (A), but for inosine.

745

746 **Figure S13. m^6A and Ψ stoichiometry distributions in different versions of Dorado.**

747 Stoichiometry distributions (all stoichiometries – left, ≥ 10% stoichiometries – right) in a line graph format
748 for **(A)** all-context m^6A detected in genome-aligned DMSO-treated NHDF dataset, with Dorado versions
749 0.7.0, 0.8.0 and 0.9.0 colored in gradients of red, darkness increasing with the later versions. **(B)** DRACH-
750 context m^6A detected with Dorado versions 0.5.0, 0.6.0, 0.8.0 and 0.9.0, colored in gradients of red,
751 darkness increasing with the later versions. **(C)** Ψ detected with Dorado versions 0.7.0, 0.8.0 and 0.9.0,
752 colored in gradients of blue, darkness increasing with the later versions. All the datasets were filtered for
753 ≥ 20 reads.

754

755 **Figure S14. Comparison of modification sites detected with different versions of Dorado.**

756 Upset plots showing unique and shared detected modifications (filtered for ≥ 20 reads, ≥ 10%
757 stoichiometry) in modification probability-unfiltered and filtered genome-aligned DMSO-treated NHDF
758 data. The versions are ordered on the plot by the number of detected sites, in ascending order (on the
759 left). Results are shown for the following models and versions: **(A)** all-context m^6A detected with v.0.7.0,
760 v0.8.0 and v0.9.0, **(B)** DRACH-context m^6A detected with v.0.5.0, v0.6.0, v0.8.0 and v0.9.0, **(C)** Ψ detected
761 with v.0.7.0, v0.8.0 and v0.9.0. **(D)** Nucleotide frequencies in the 5-base motif of Ψ sites for uniquely
762 v0.7.0-, v0.8.0-0.9.0- and v0.7.0-0.8.0-0.9.0-detected Ψ sites. Position 0 marks the modification site.

763

764 **Figure S15. Comparison of m^6A sites detected by Dorado and GLORI-Sequencing in NHDFs and HD10.6s.**

765 **(A)** Upset plot showing the number of m^6A sites detected by GLORI and Dorado in NHDFs and their
766 overlaps. Dorado sites represent high-confidence m^6A sites obtained after filtering DRACH-model detected
767 data. Similarly, only high-confidence m^6A sites detected by GLORI are shown, but were additionally filtered
768 for DRACH sites. **(B)** Stoichiometry distributions are plotted from genome-aligned DMSO- or STM2457-
769 treated HD10.6 datasets using the indicated Dorado models. **(C)** Stoichiometry distributions from GLORI
770 performed on DMSO- or STM2457-treated NHDFs are plotted. In addition, m^6A sites were filtered for
771 DRACH motifs (right plot). **(D)** Stoichiometry distributions are plotted from transcriptome-aligned DMSO-
772 or STM2457-treated NHDF datasets using the indicated Dorado models. **(E)** Stoichiometry distributions
773 are plotted from transcriptome-aligned DMSO- or STM2457-treated HD10.6 datasets using the indicated
774 Dorado models.

775

776 **Figure S16. m^6A (all-context and DRACH-context) motif logos for modification probability-unfiltered**
777 **NHDF and HD10.6 data.**

778 **(A)** 5-mer motif logo (showing nucleotide frequences at each position) of m^6A-modified sites (position 0
779 = modified A) from genome-aligned DMSO-treated NHDFs (all-context – left, DRACH-context – right). All
780 the datasets were filtered for ≥ 20 coverage & ≥ 10% stoichiometry. **(B)** Same as (A), but for HD10.6 cells.

781

782 **Figure S17. Metagene plots showing m^6A (all-context and DRACH-context) distribution profiles in NHDF**
783 **and HD10.6 data filtered at different stoichiometries.**

784 **(A)** Metagene plots showing the density of m^6A (all-context – dark red, DRACH-context – bright red) sites
785 across all annotated poly(A) transcripts (upper plot) and monoexonic poly(A) transcripts (lower plot) from
786 genome-aligned DMSO-treated NHDF datasets filtered with 0.98 modification probability, ≥ 20 coverage
787 and ≥ 10% stoichiometry. **(B)** The density of all-context m^6A sites (0.98 modification probability-unfiltered
788 – left, filtered – right) is plotted across all annotated poly(A) transcripts of NHDF (dark red) and HD10.6
789 (bright red) datasets, filtered with different stoichiometries (descending: all stoichiometries, ≥ 50%, ≥ 10%
790 and < 10%. All datasets were filtered additionally for ≥ 20 coverage & ≥ 10% stoichiometry. **(C)** Same as
791 (A), but for DRACH-context m^6A sites.

Figure 1

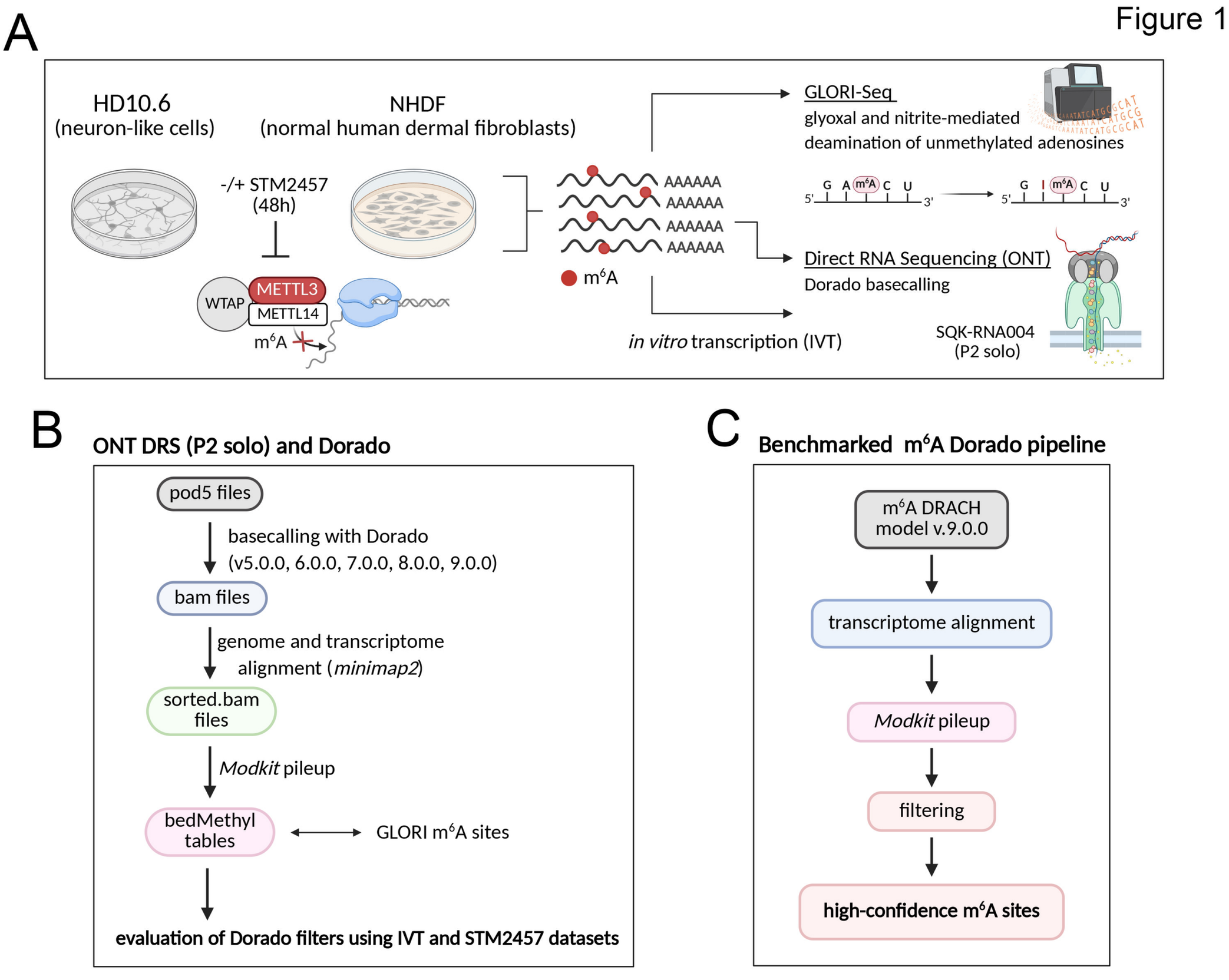

A
HD10.6
(neuron-like cells)
NHDF
(normal human dermal fibroblasts)
-/+ STM2457
(48h)
WTAP
METTL3
METTL14
m6A
AAAAAA
AAAAAA
AAAAAA
AAAAAA
m6A
GLORI-Seq
glyoxal and nitrite-mediated
deamination of unmethylated adenosines
G A m6A C U
5' 3'
G I m6A C U
5' 3'
Direct RNA Sequencing (ONT)
Dorado basecalling
in vitro transcription (IVT)
SQK-RNA004
(P2 solo)
B
ONT DRS (P2 solo) and Dorado
pod5 files
basecalling with Dorado
(v5.0.0, 6.0.0, 7.0.0, 8.0.0, 9.0.0)
bam files
genome and transcriptome
alignment (minimap2)
sorted.bam
files
Modkit pileup
bedMethyl
tables
GLORI m6A sites
evaluation of Dorado filters using IVT and STM2457 datasets
C
Benchmarked m6A Dorado pipeline
m6A DRACH
model v.9.0.0
transcriptome alignment
Modkit pileup
filtering
high-confidence m6A sites

Figure 2

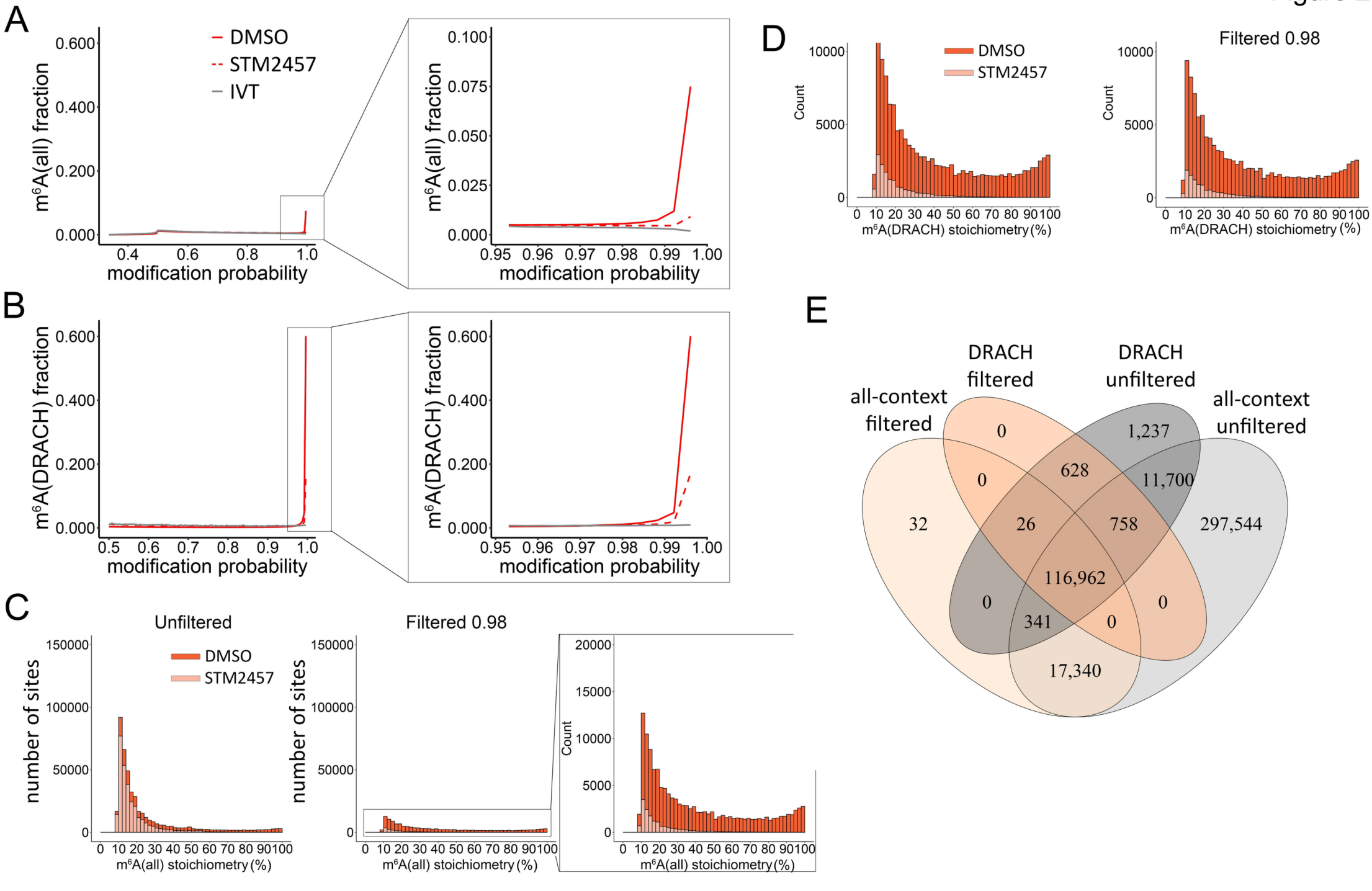

A
DMSO
STM2457
IVT
m6A(all) fraction
modification probability
B
m6A(DRACH) fraction
modification probability
C
Unfiltered
Filtered 0.98
DMSO
STM2457
number of sites
Count
m6A(all) stoichiometry (%)
D
Filtered 0.98
Count
m6A(DRACH) stoichiometry (%)
E
all-context filtered
DRACH filtered
DRACH unfiltered
all-context unfiltered
0
1,237
628
11,700
0
32
26
758
297,544
116,962
0
0
341
0
17,340

Figure 3

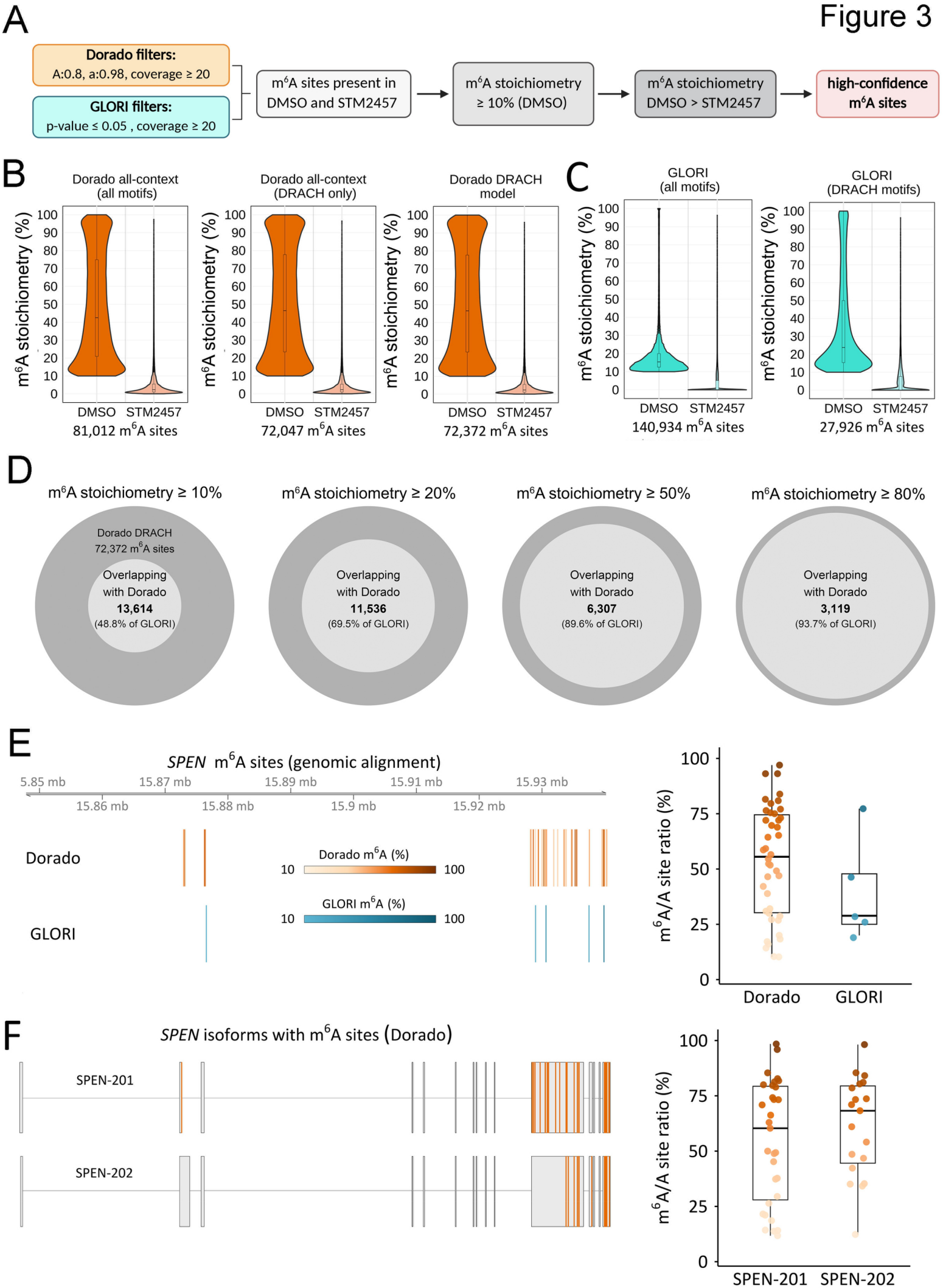

A
Dorado filters:
A:0.8, a:0.98, coverage ≥ 20
GLORI filters:
p-value ≤ 0.05 , coverage ≥ 20
m6A sites present in DMSO and STM2457
m6A stoichiometry ≥ 10% (DMSO)
m6A stoichiometry DMSO > STM2457
high-confidence m6A sites
B
Dorado all-context (all motifs)
Dorado all-context (DRACH only)
Dorado DRACH model
m6A stoichiometry (%)
DMSO STM2457
81,012 m6A sites
72,047 m6A sites
72,372 m6A sites
C
GLORI (all motifs)
GLORI (DRACH motifs)
140,934 m6A sites
27,926 m6A sites
D
m6A stoichiometry ≥ 10%
m6A stoichiometry ≥ 20%
m6A stoichiometry ≥ 50%
m6A stoichiometry ≥ 80%
Dorado DRACH 72,372 m6A sites
Overlapping with Dorado 13,614 (48.8% of GLORI)
Overlapping with Dorado 11,536 (69.5% of GLORI)
Overlapping with Dorado 6,307 (89.6% of GLORI)
Overlapping with Dorado 3,119 (93.7% of GLORI)
E
SPEN m6A sites (genomic alignment)
5.85 mb 15.87 mb 15.89 mb 15.91 mb 15.93 mb
15.86 mb 15.88 mb 15.9 mb 15.92 mb
Dorado
GLORI
Dorado m6A (%)
GLORI m6A (%)
m6A/A site ratio (%)
Dorado GLORI
F
SPEN isoforms with m6A sites (Dorado)
SPEN-201
SPEN-202
m6A/A site ratio (%)
SPEN-201 SPEN-202

Figure 4

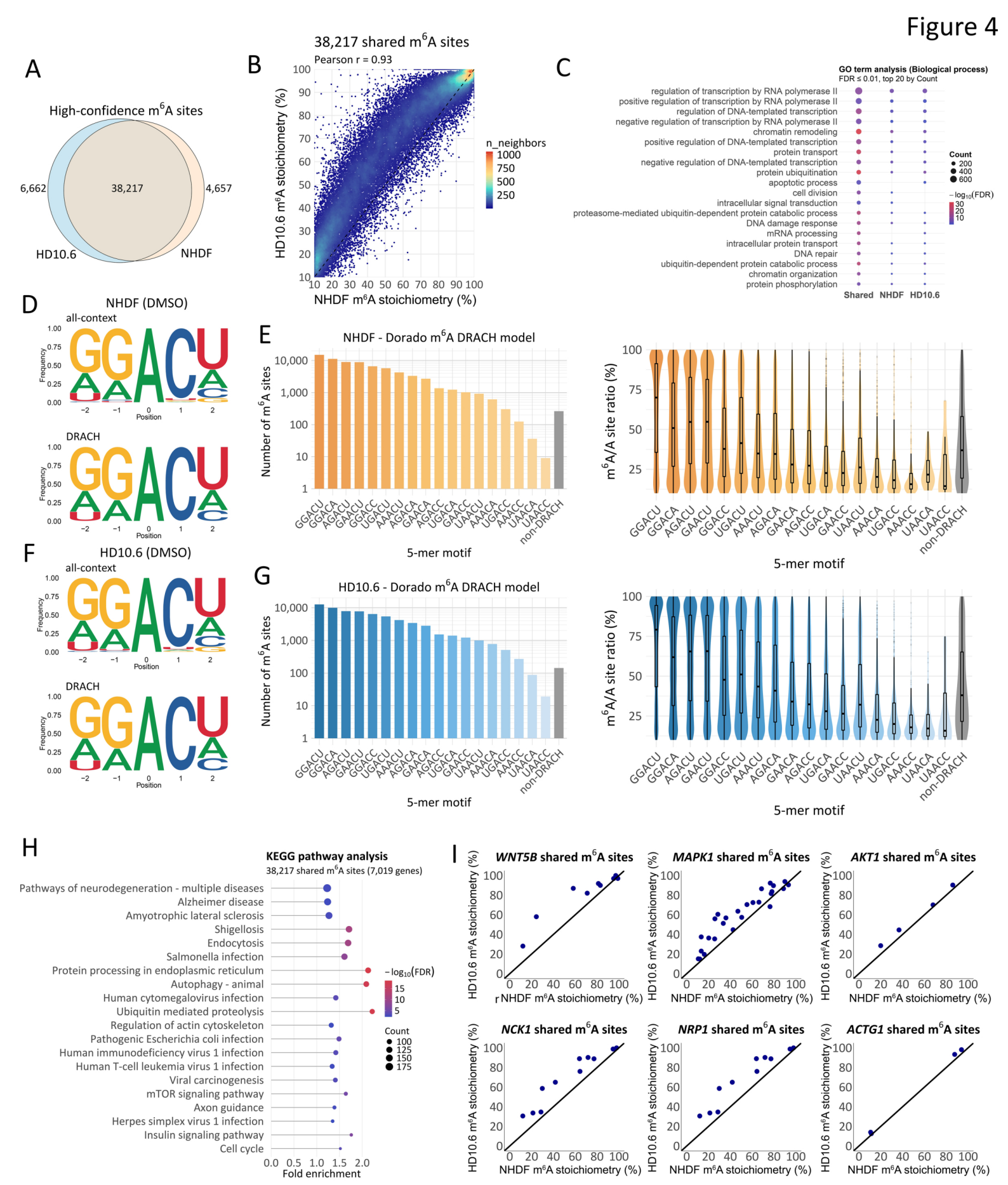

A
High-confidence m6A sites
6,662
38,217
4,657
HD10.6
NHDF
B
38,217 shared m6A sites
Pearson r = 0.93
HD10.6 m6A stoichiometry (%)
NHDF m6A stoichiometry (%)
n_neighbors
1000
750
500
250
C
GO term analysis (Biological process)
FDR ≤ 0.01, top 20 by Count
regulation of transcription by RNA polymerase II
positive regulation of transcription by RNA polymerase II
regulation of DNA-templated transcription
negative regulation of transcription by RNA polymerase II
chromatin remodeling
positive regulation of DNA-templated transcription
protein transport
negative regulation of DNA-templated transcription
protein ubiquitination
apoptotic process
cell division
intracellular signal transduction
proteasome-mediated ubiquitin-dependent protein catabolic process
DNA damage response
mRNA processing
intracellular protein transport
DNA repair
ubiquitin-dependent protein catabolic process
chromatin organization
protein phosphorylation
Shared
NHDF
HD10.6
Count
200
400
600
−log10(FDR)
30
20
10
D
NHDF (DMSO)
all-context
DRACH
Frequency
Position
E
NHDF - Dorado m6A DRACH model
Number of m6A sites
5-mer motif
m6A/A site ratio (%)
GGACU
GGACA
AGACU
GAACU
GGACC
UGACU
AAACU
AGACA
GAACA
AGACC
UGACA
GAACC
UAACU
AAACA
UGACC
AAACC
UAACA
UAACC
non-DRACH
F
HD10.6 (DMSO)
all-context
DRACH
G
HD10.6 - Dorado m6A DRACH model
H
KEGG pathway analysis
38,217 shared m6A sites (7,019 genes)
Pathways of neurodegeneration - multiple diseases
Alzheimer disease
Amyotrophic lateral sclerosis
Shigellosis
Endocytosis
Salmonella infection
Protein processing in endoplasmic reticulum
Autophagy - animal
Human cytomegalovirus infection
Ubiquitin mediated proteolysis
Regulation of actin cytoskeleton
Pathogenic Escherichia coli infection
Human immunodeficiency virus 1 infection
Human T-cell leukemia virus 1 infection
Viral carcinogenesis
mTOR signaling pathway
Axon guidance
Herpes simplex virus 1 infection
Insulin signaling pathway
Cell cycle
Fold enrichment
−log10(FDR)
15
10
5
Count
100
125
150
175
I
WNT5B shared m6A sites
MAPK1 shared m6A sites
AKT1 shared m6A sites
NCK1 shared m6A sites
NRP1 shared m6A sites
ACTG1 shared m6A sites
HD10.6 m6A stoichiometry (%)
NHDF m6A stoichiometry (%)

Figure S1

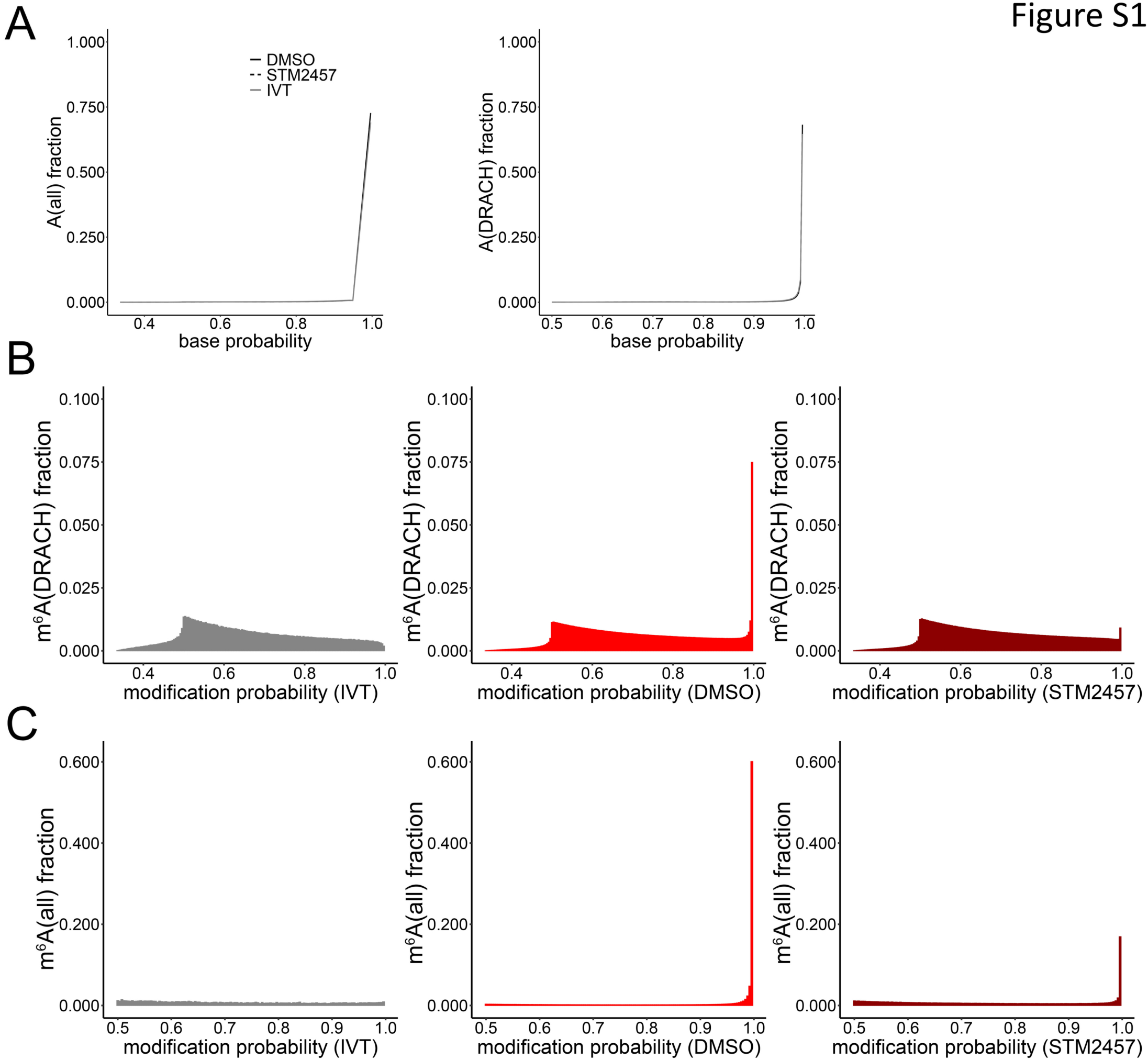

A
DMSO
STM2457
IVT
A(all) fraction
A(DRACH) fraction
base probability
B
m6A(DRACH) fraction
modification probability (IVT)
modification probability (DMSO)
modification probability (STM2457)
C
m6A(all) fraction
modification probability (IVT)
modification probability (DMSO)
modification probability (STM2457)

Figure S2

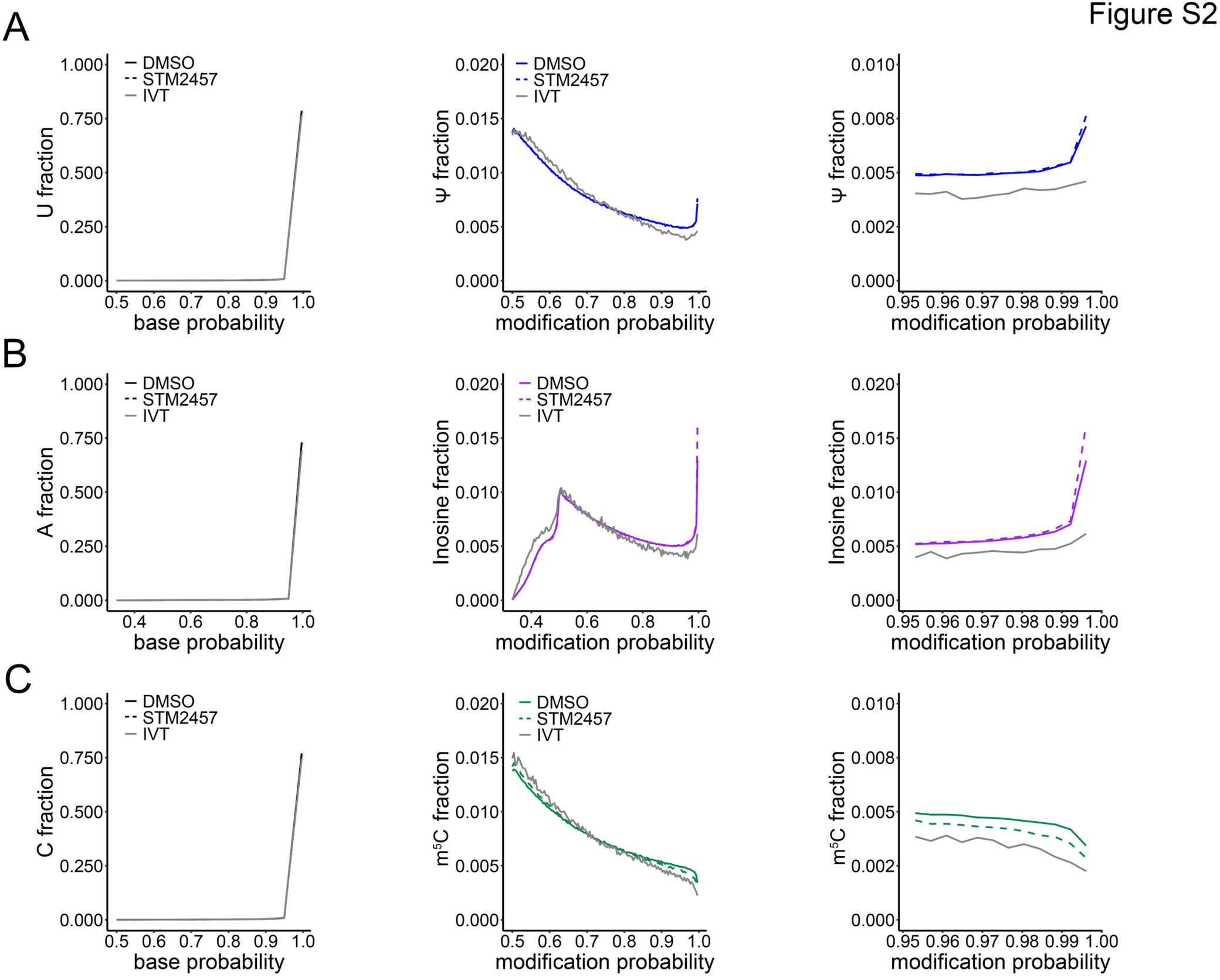
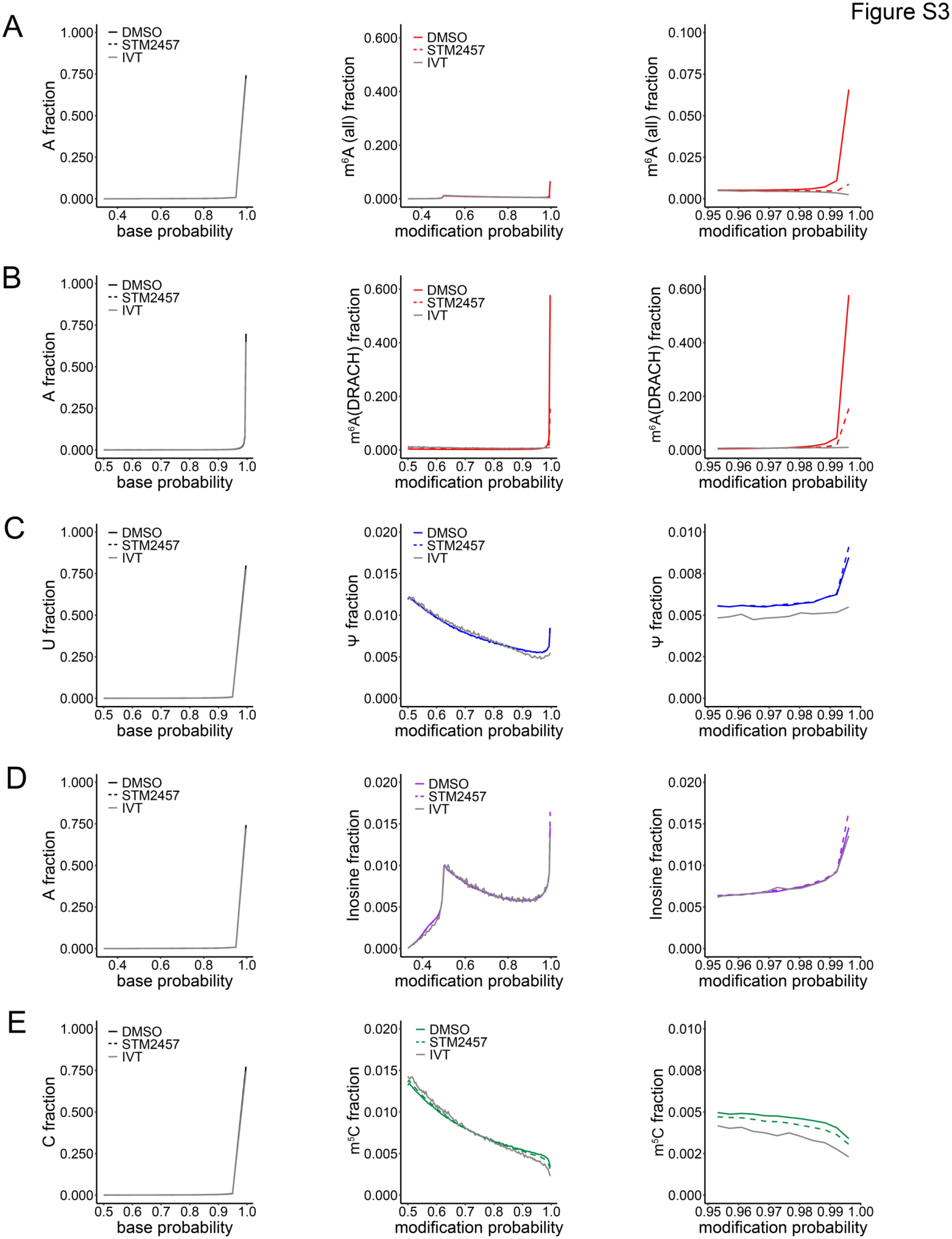

A
DMSO
STM2457
IVT
U fraction
base probability
ψ fraction
modification probability
B
A fraction
Inosine fraction
C
C fraction
m5C fraction

Figure S3


A
DMSO
STM2457
IVT
A fraction
base probability
m6A (all) fraction
modification probability
B
m6A(DRACH) fraction
C
U fraction
Ψ fraction
D
Inosine fraction
E
C fraction
m5C fraction

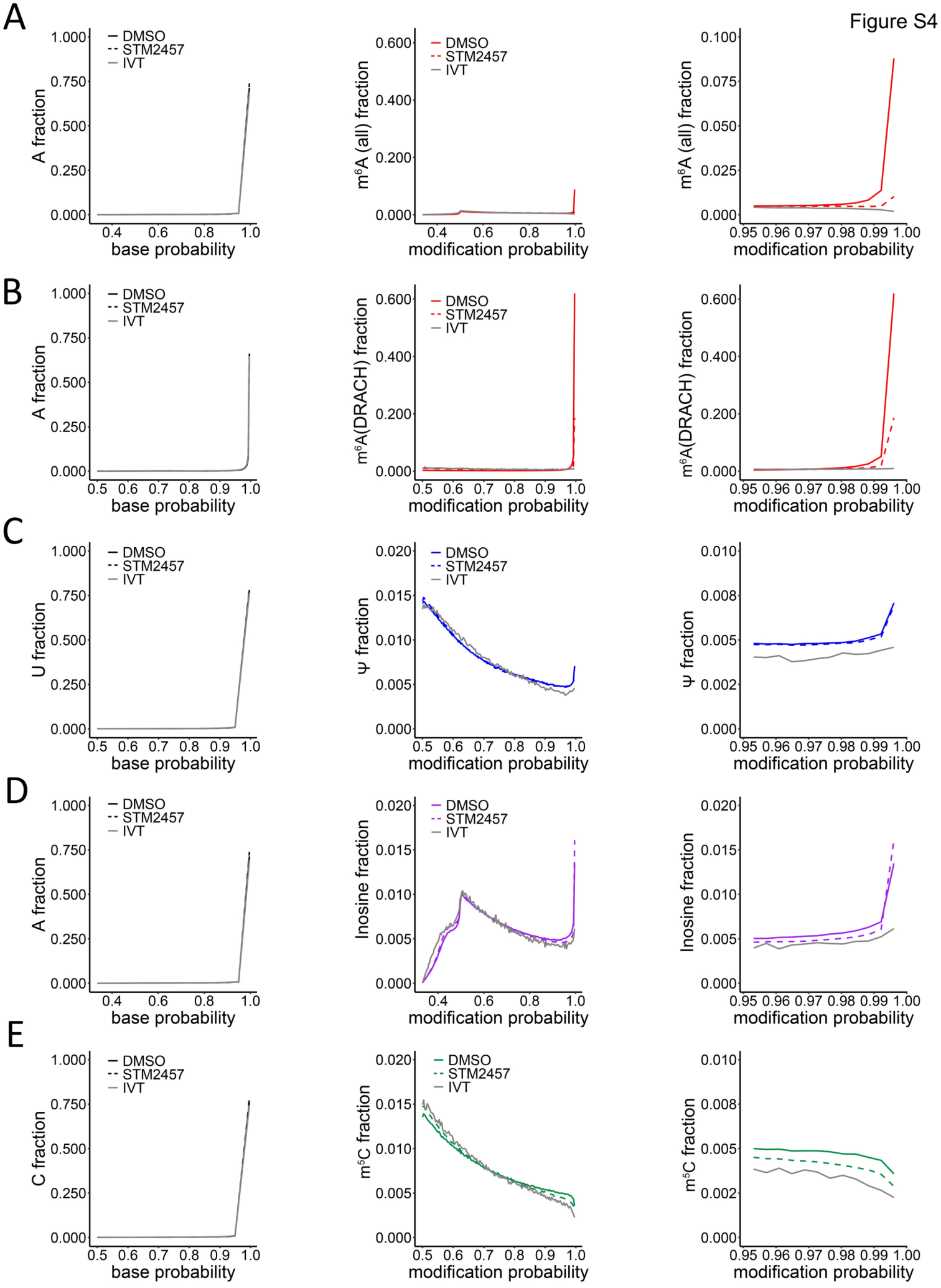


Figure S4
A
B
C
D
E
DMSO
STM2457
IVT
A fraction
m6A (all) fraction
m6A(DRACH) fraction
U fraction
Ψ fraction
Inosine fraction
C fraction
m5C fraction
base probability
modification probability

Figure S5

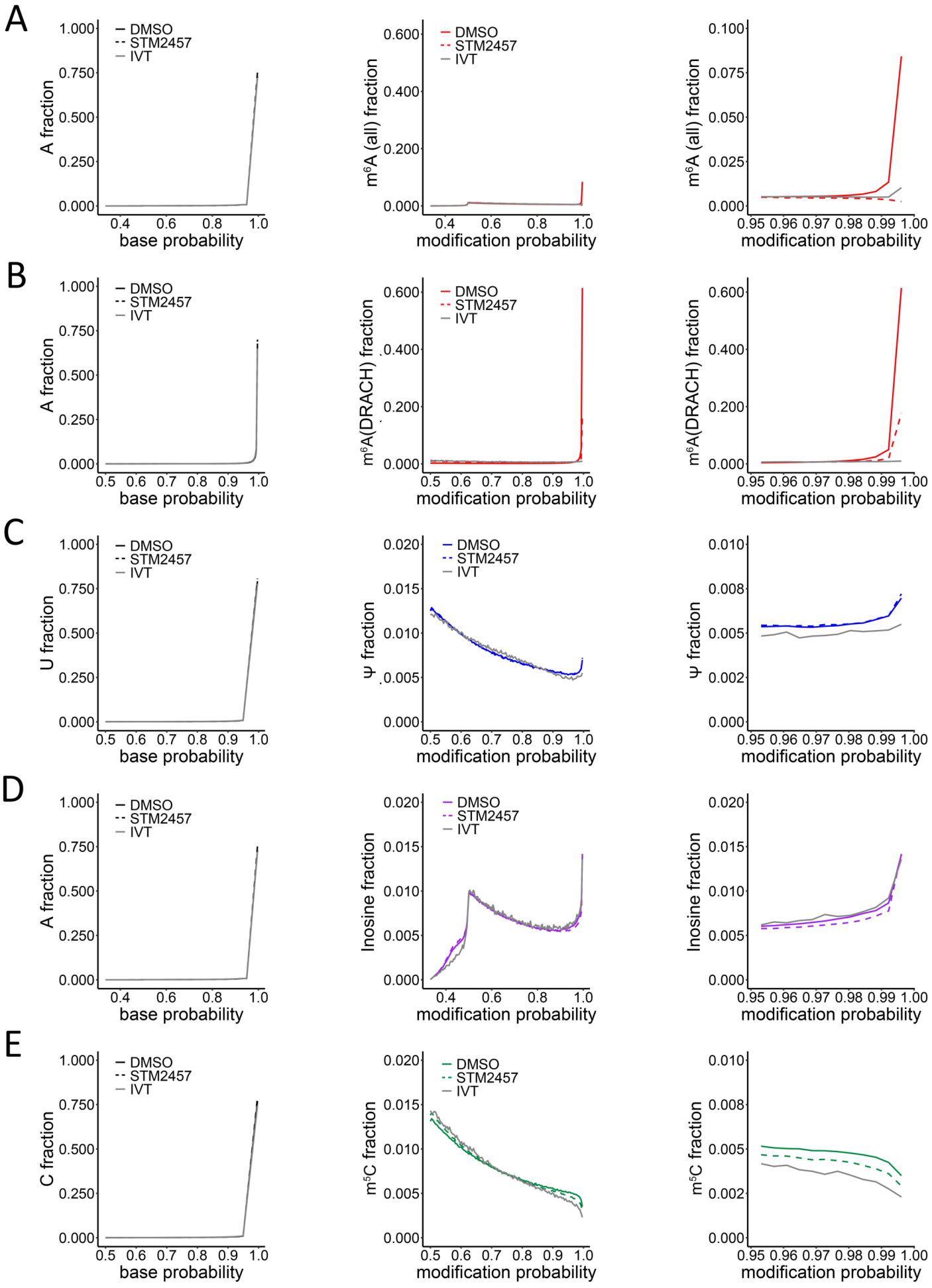

A
B
C
D
E
DMSO
STM2457
IVT
A fraction
U fraction
C fraction
m6A (all) fraction
m6A(DRACH) fraction
Ψ fraction
Inosine fraction
m5C fraction
base probability
modification probability

Figure S6

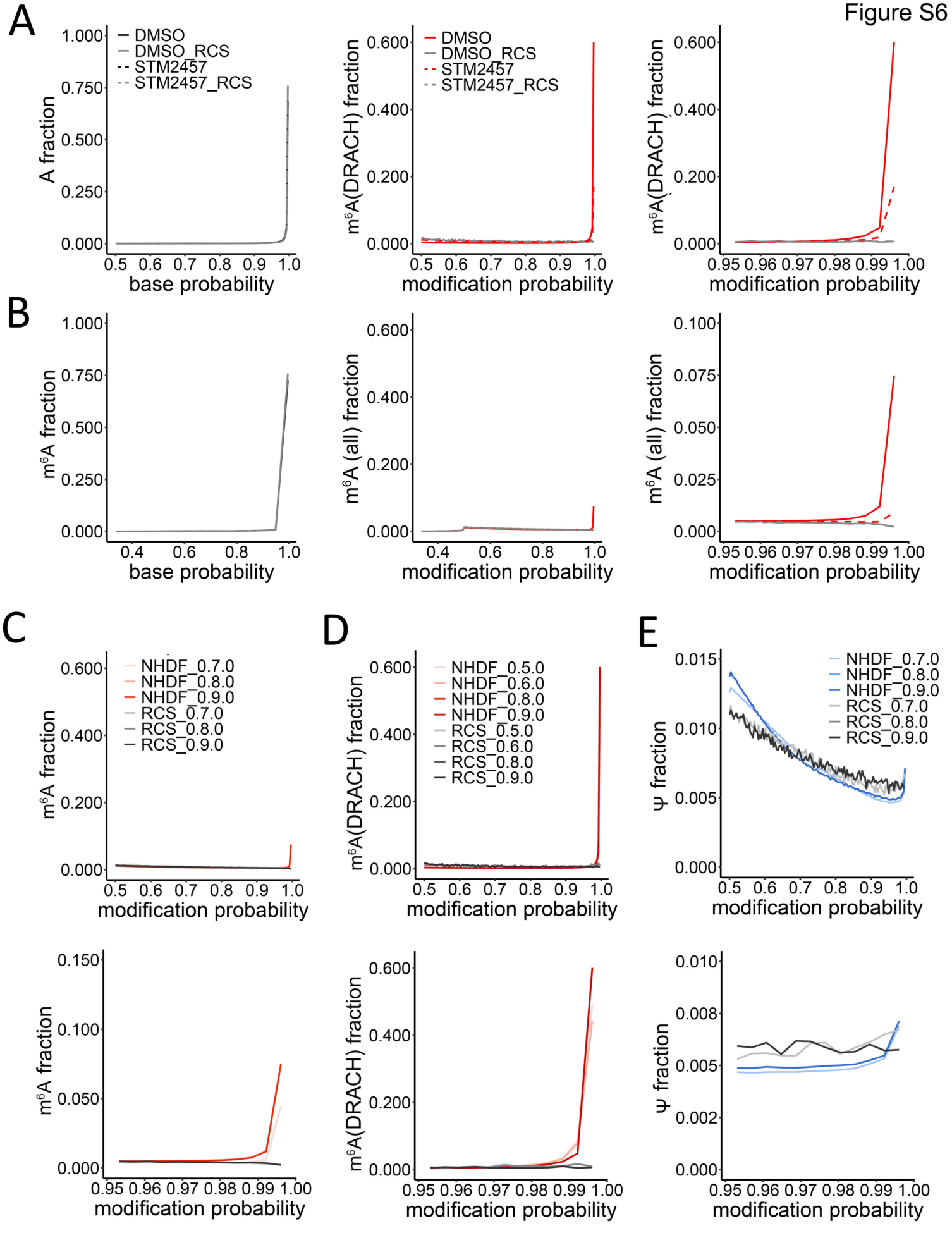

A
DMSO
DMSO_RCS
STM2457
STM2457_RCS
A fraction
base probability
m6A(DRACH) fraction
modification probability
B
m6A fraction
m6A (all) fraction
C
NHDF_0.7.0
NHDF_0.8.0
NHDF_0.9.0
RCS_0.7.0
RCS_0.8.0
RCS_0.9.0
D
NHDF_0.5.0
NHDF_0.6.0
NHDF_0.8.0
NHDF_0.9.0
RCS_0.5.0
RCS_0.6.0
RCS_0.8.0
RCS_0.9.0
E
ψ fraction

Figure S7

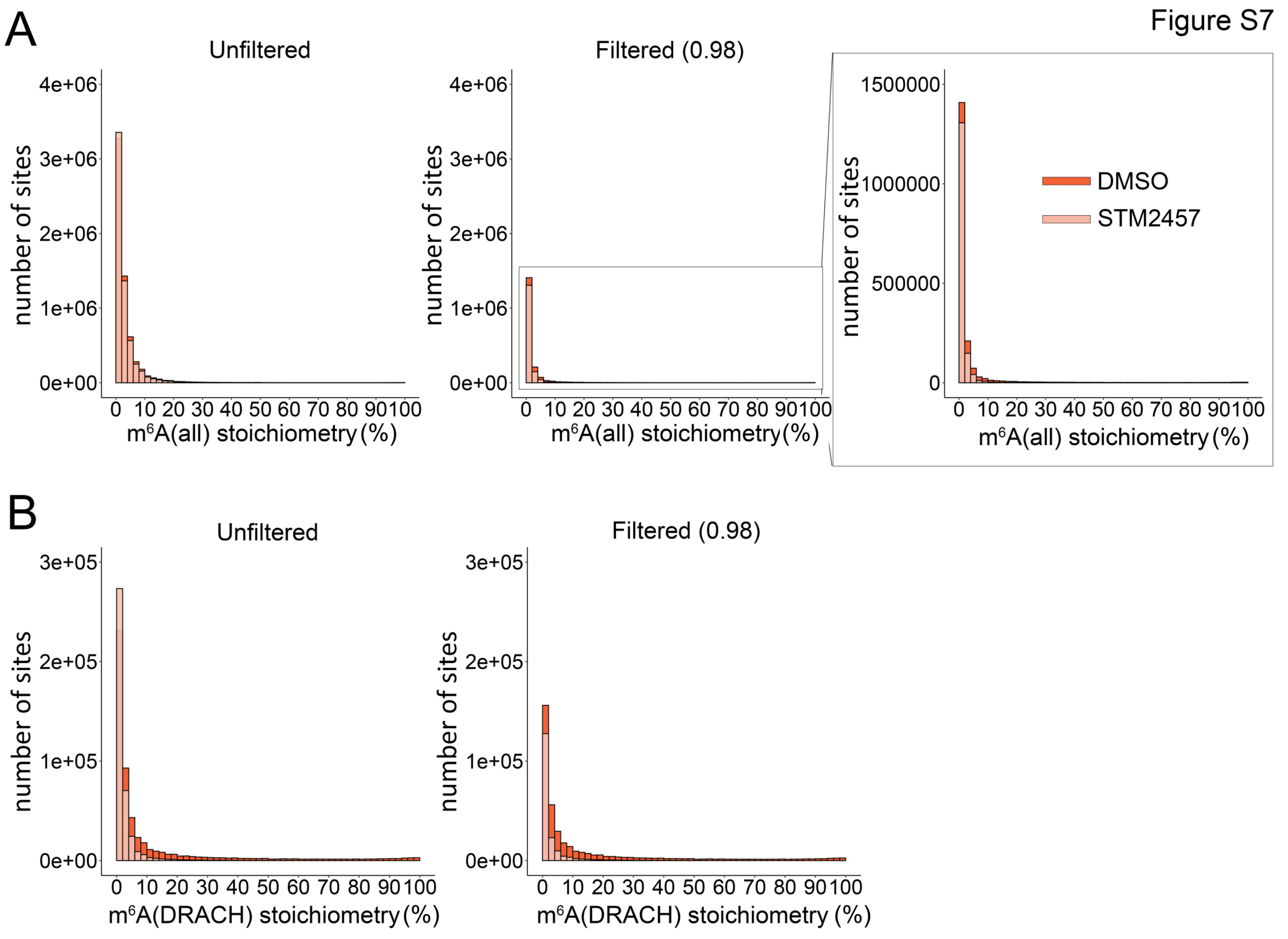

A
Unfiltered
Filtered (0.98)
number of sites
4e+06
3e+06
2e+06
1e+06
0e+00
1500000
1000000
500000
0
0 10 20 30 40 50 60 70 80 90 100
m6A(all) stoichiometry (%)
DMSO
STM2457
B
3e+05
2e+05
1e+05
0e+00
m6A(DRACH) stoichiometry (%)

Figure S8

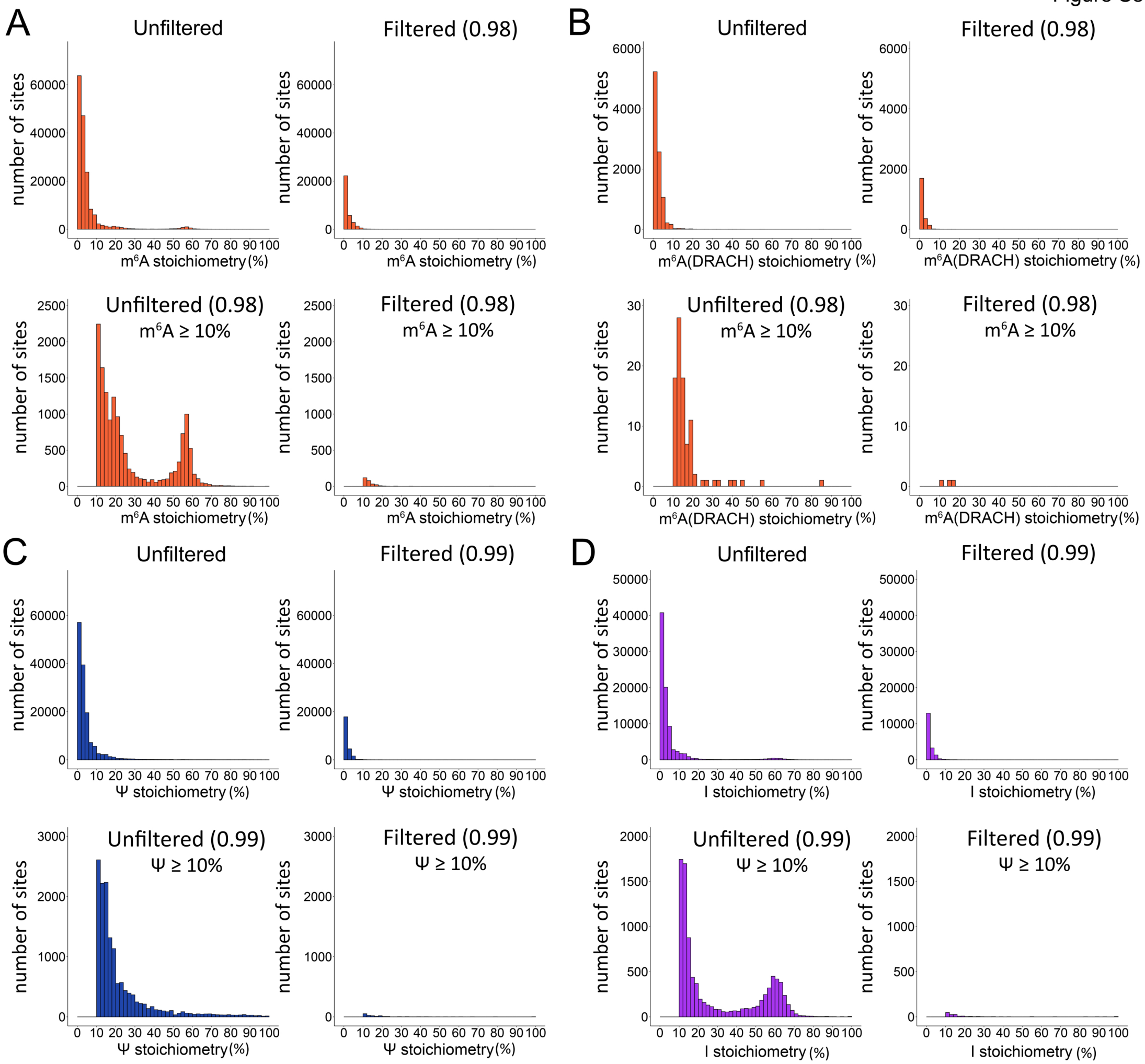
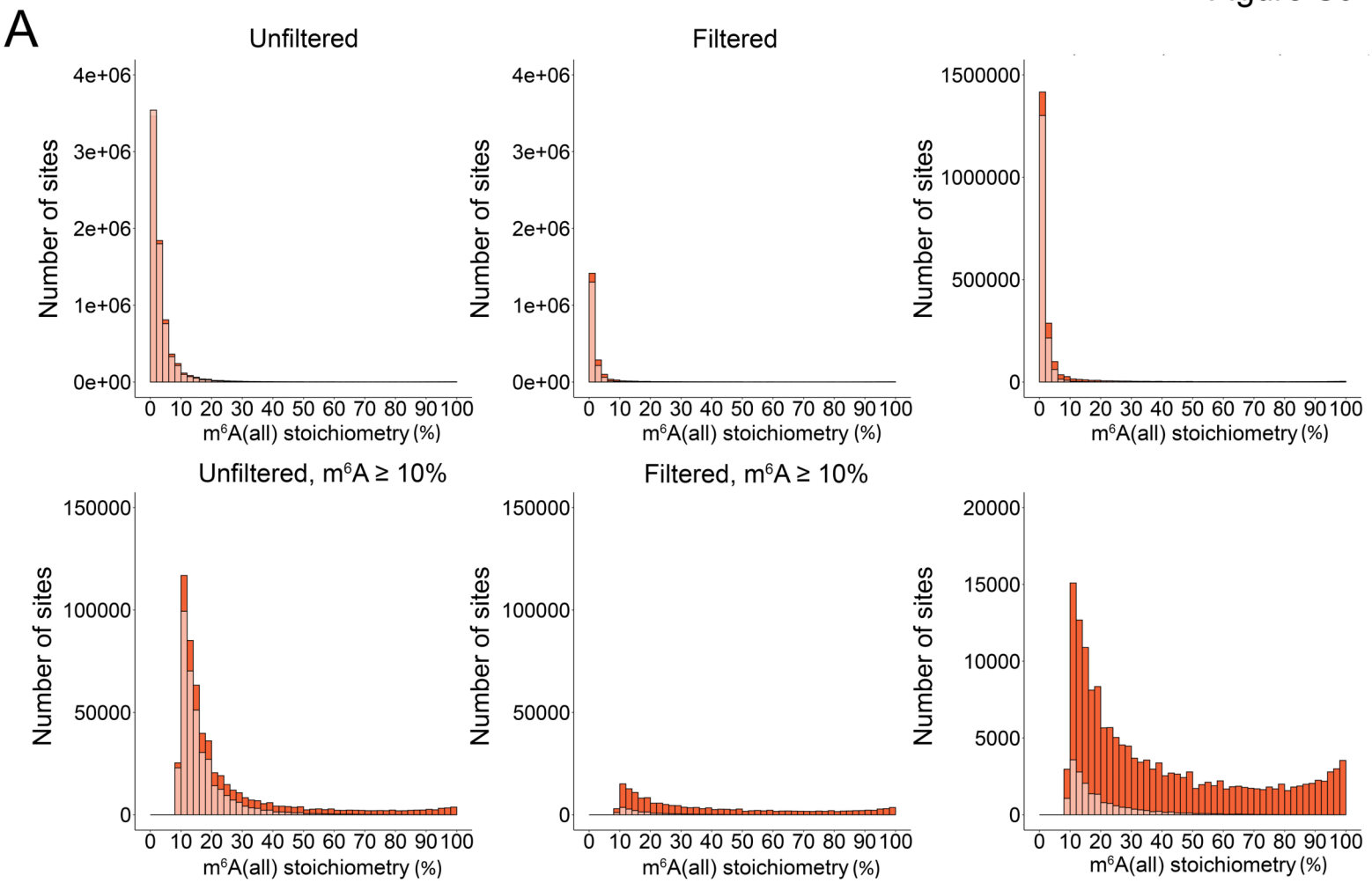

A
Unfiltered
Filtered (0.98)
Unfiltered (0.98)
m6A ≥ 10%
Filtered (0.98)
m6A ≥ 10%
number of sites
m6A stoichiometry (%)
B
Unfiltered
Filtered (0.98)
Unfiltered (0.98)
m6A ≥ 10%
Filtered (0.98)
m6A ≥ 10%
number of sites
m6A(DRACH) stoichiometry (%)
C
Unfiltered
Filtered (0.99)
Unfiltered (0.99)
Ψ ≥ 10%
Filtered (0.99)
Ψ ≥ 10%
number of sites
Ψ stoichiometry (%)
D
Unfiltered
Filtered (0.99)
Unfiltered (0.99)
Ψ ≥ 10%
Filtered (0.99)
Ψ ≥ 10%
number of sites
I stoichiometry (%)

Figure S9


A
Unfiltered
Filtered
Number of sites
4e+06
3e+06
2e+06
1e+06
0e+00
1500000
1000000
500000
0
0 10 20 30 40 50 60 70 80 90 100
m6A(all) stoichiometry (%)
Unfiltered, m6A ≥ 10%
Filtered, m6A ≥ 10%
150000
100000
50000
20000
15000
10000
5000

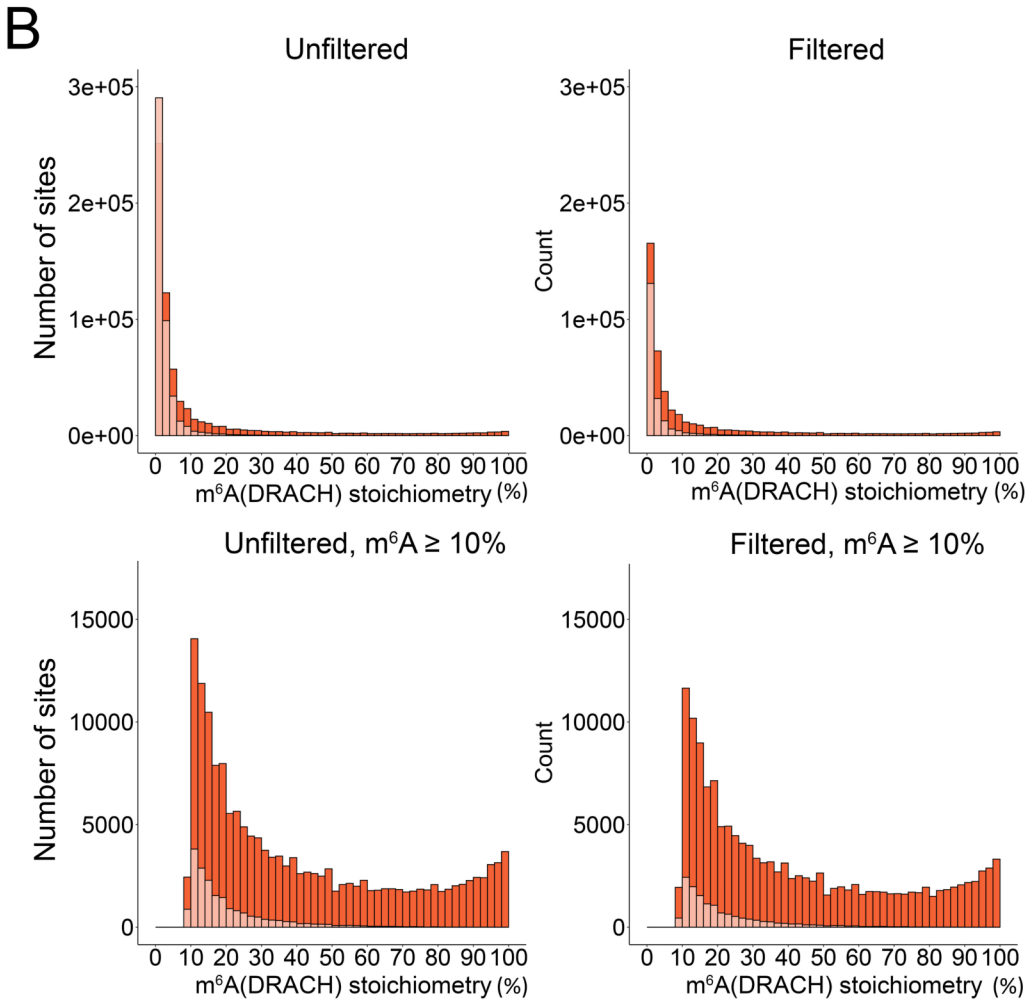
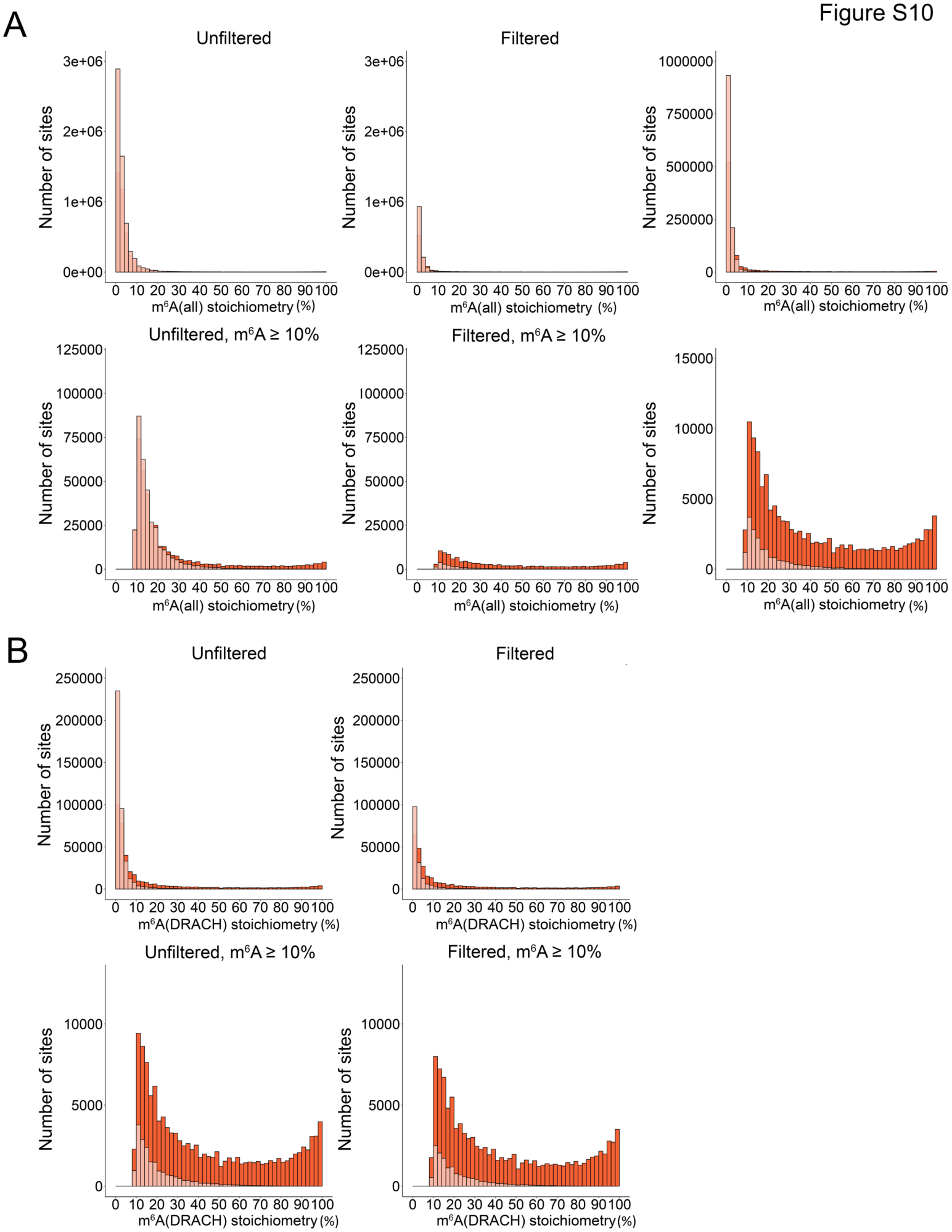
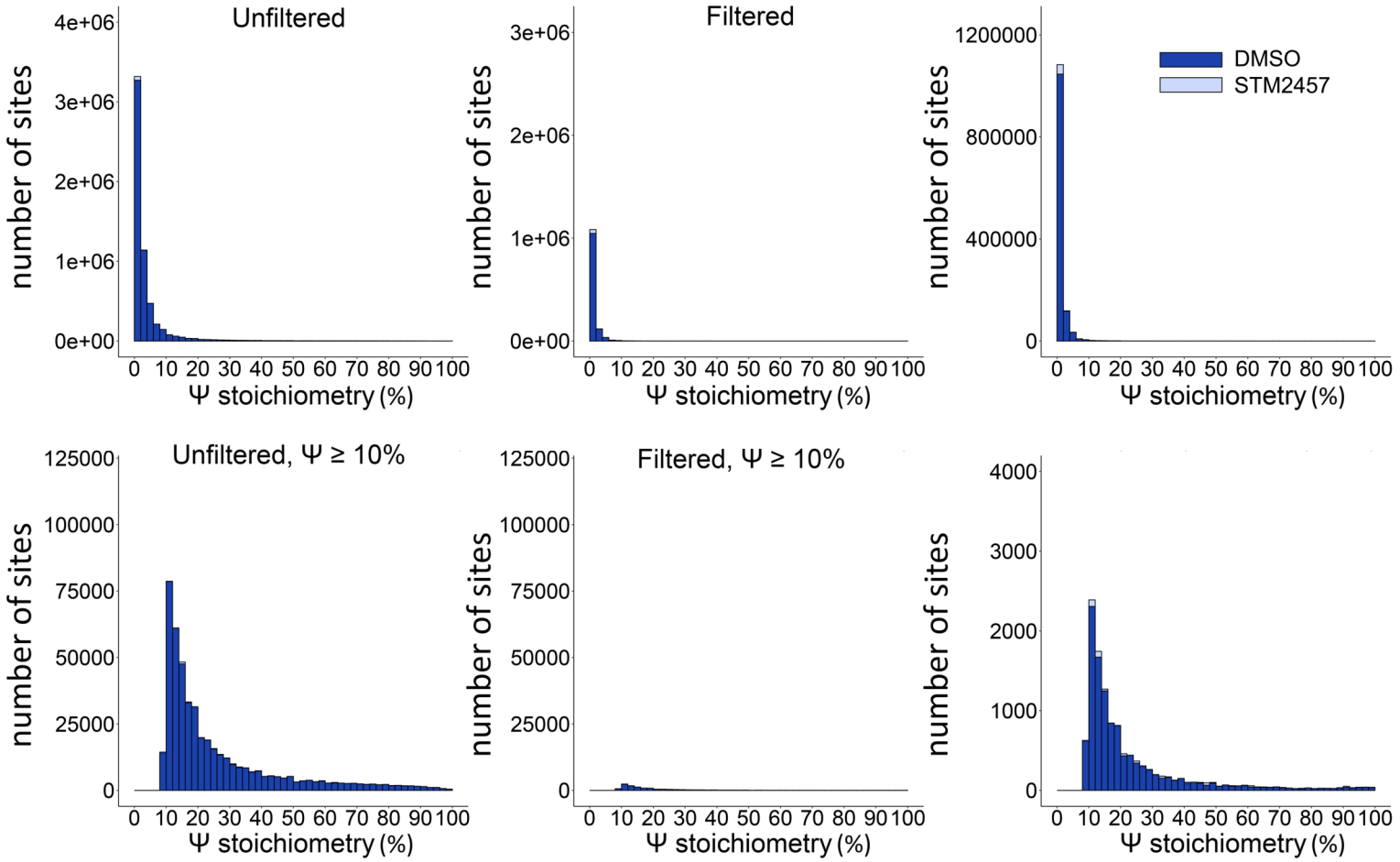

B
Unfiltered
Filtered
Number of sites
Count
3e+05
2e+05
1e+05
0e+00
15000
10000
5000
0
0 10 20 30 40 50 60 70 80 90 100
m6A(DRACH) stoichiometry (%)
Unfiltered, m6A ≥ 10%
Filtered, m6A ≥ 10%

Figure S10


A
Unfiltered
Filtered
Unfiltered, m6A ≥ 10%
Filtered, m6A ≥ 10%
Number of sites
m6A(all) stoichiometry (%)
B
Unfiltered
Filtered
Unfiltered, m6A ≥ 10%
Filtered, m6A ≥ 10%
Number of sites
m6A(DRACH) stoichiometry (%)

A

Figure S11



B

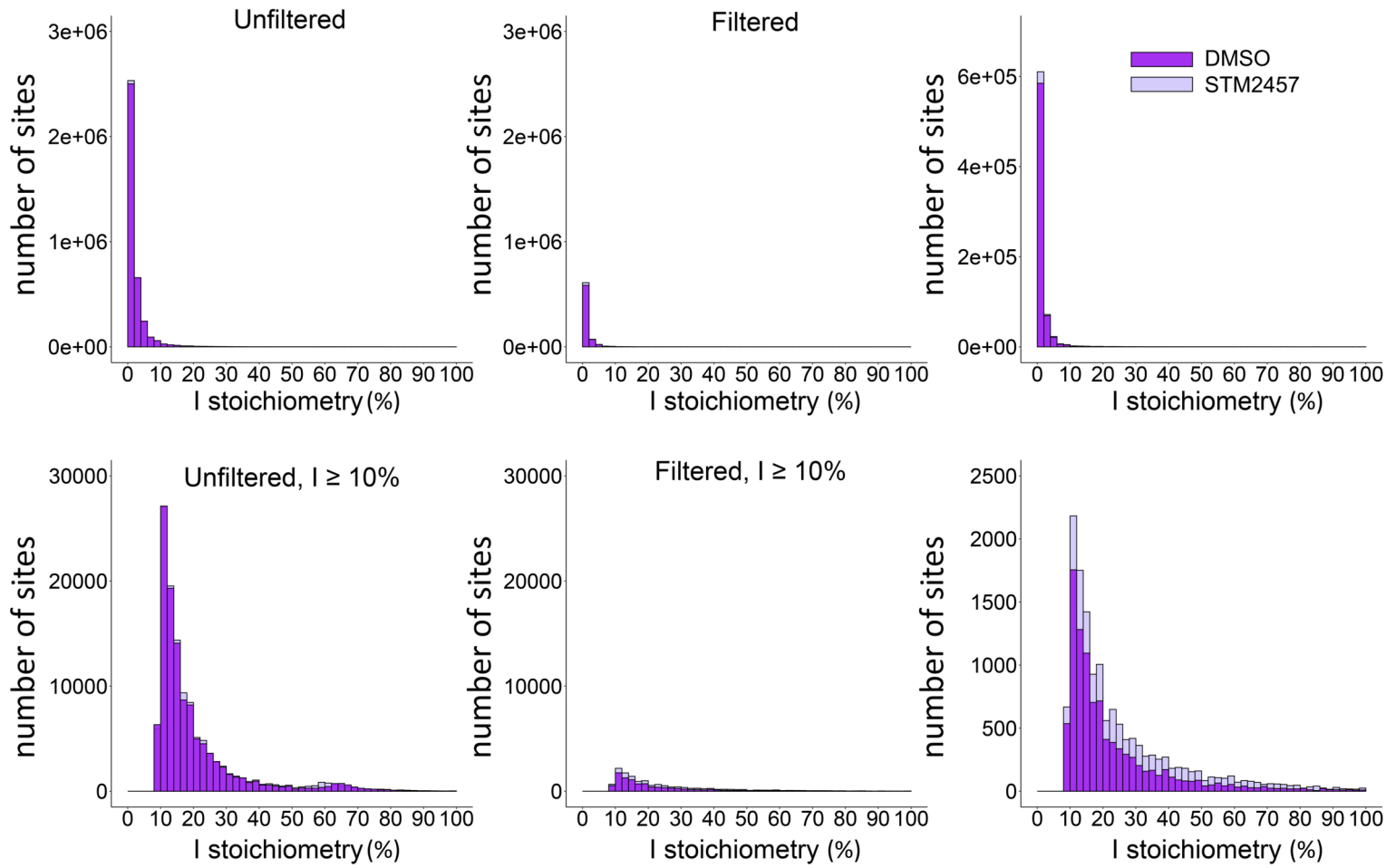


Figure S12

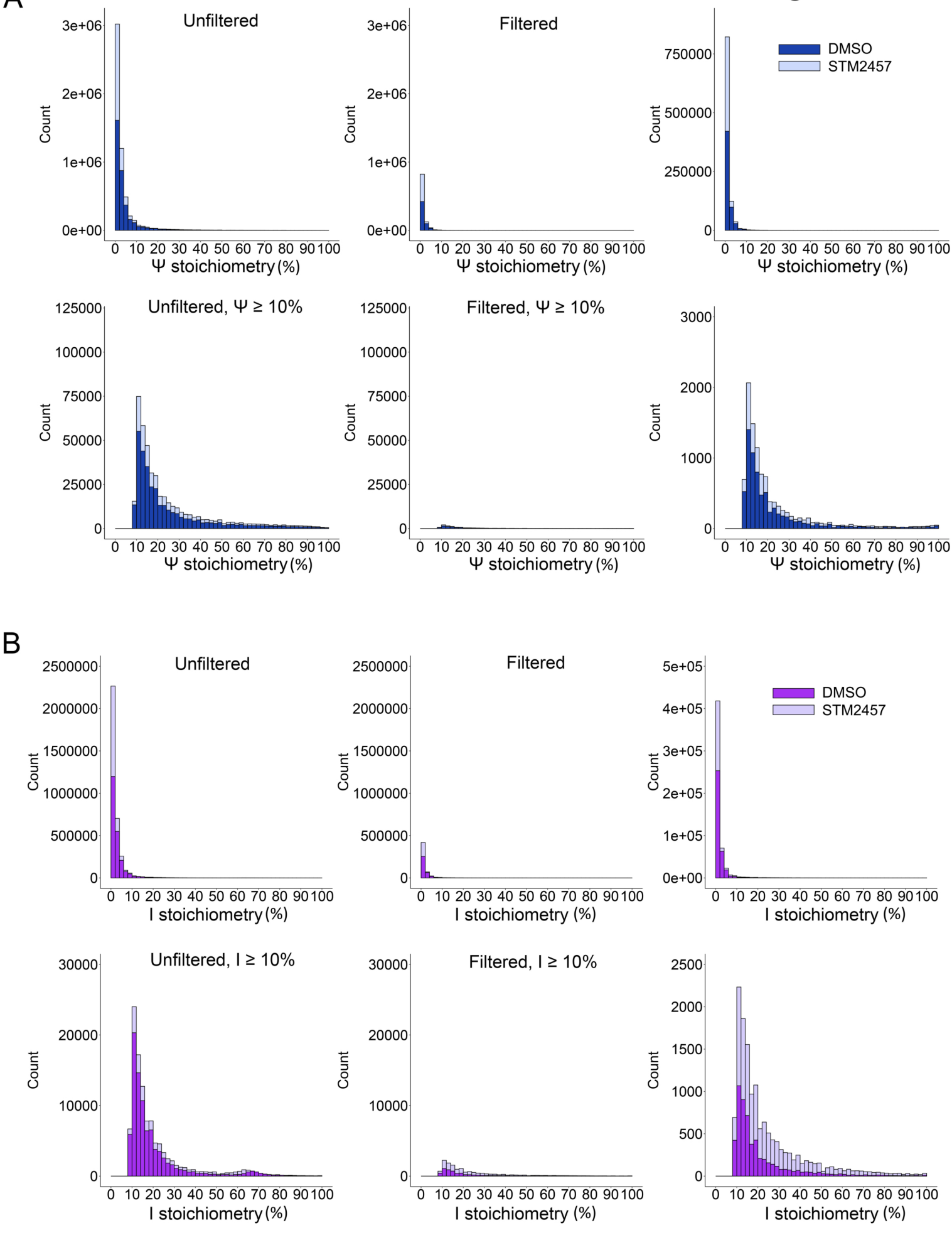

A
Unfiltered
Filtered
DMSO
STM2457
Count
Ψ stoichiometry (%)
Unfiltered, Ψ ≥ 10%
Filtered, Ψ ≥ 10%
B
Unfiltered
Filtered
DMSO
STM2457
Count
I stoichiometry (%)
Unfiltered, I ≥ 10%
Filtered, I ≥ 10%

Figure S13

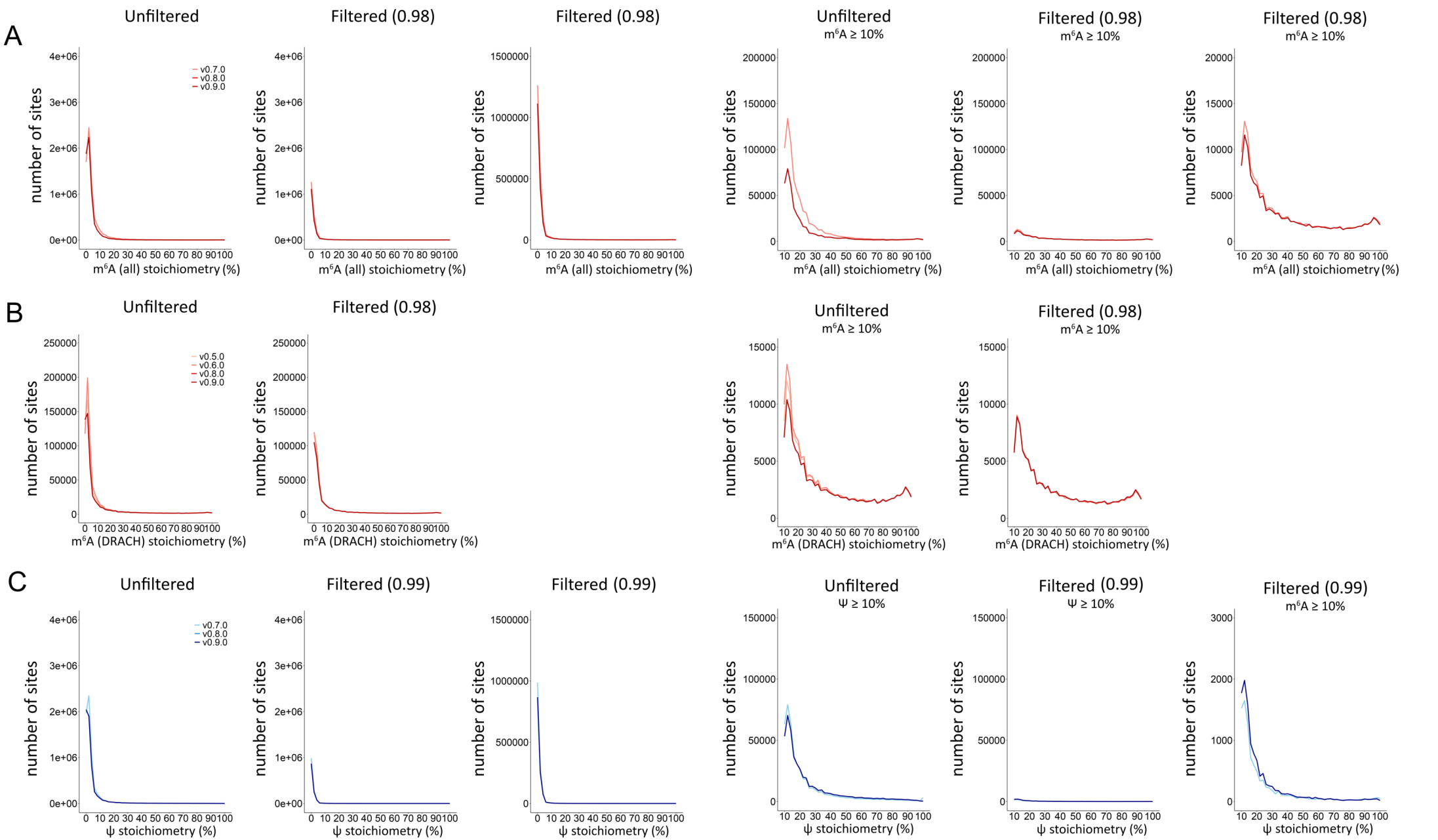

A
Unfiltered
Filtered (0.98)
Filtered (0.98)
Unfiltered
m6A ≥ 10%
Filtered (0.98)
m6A ≥ 10%
Filtered (0.98)
m6A ≥ 10%
v0.7.0
v0.8.0
v0.9.0
number of sites
m6A (all) stoichiometry (%)
B
Unfiltered
Filtered (0.98)
Unfiltered
m6A ≥ 10%
Filtered (0.98)
m6A ≥ 10%
v0.5.0
v0.6.0
v0.8.0
v0.9.0
number of sites
m6A (DRACH) stoichiometry (%)
C
Unfiltered
Filtered (0.99)
Filtered (0.99)
Unfiltered
Ψ ≥ 10%
Filtered (0.99)
Ψ ≥ 10%
Filtered (0.99)
m6A ≥ 10%
v0.7.0
v0.8.0
v0.9.0
number of sites
ψ stoichiometry (%)

Figure S14

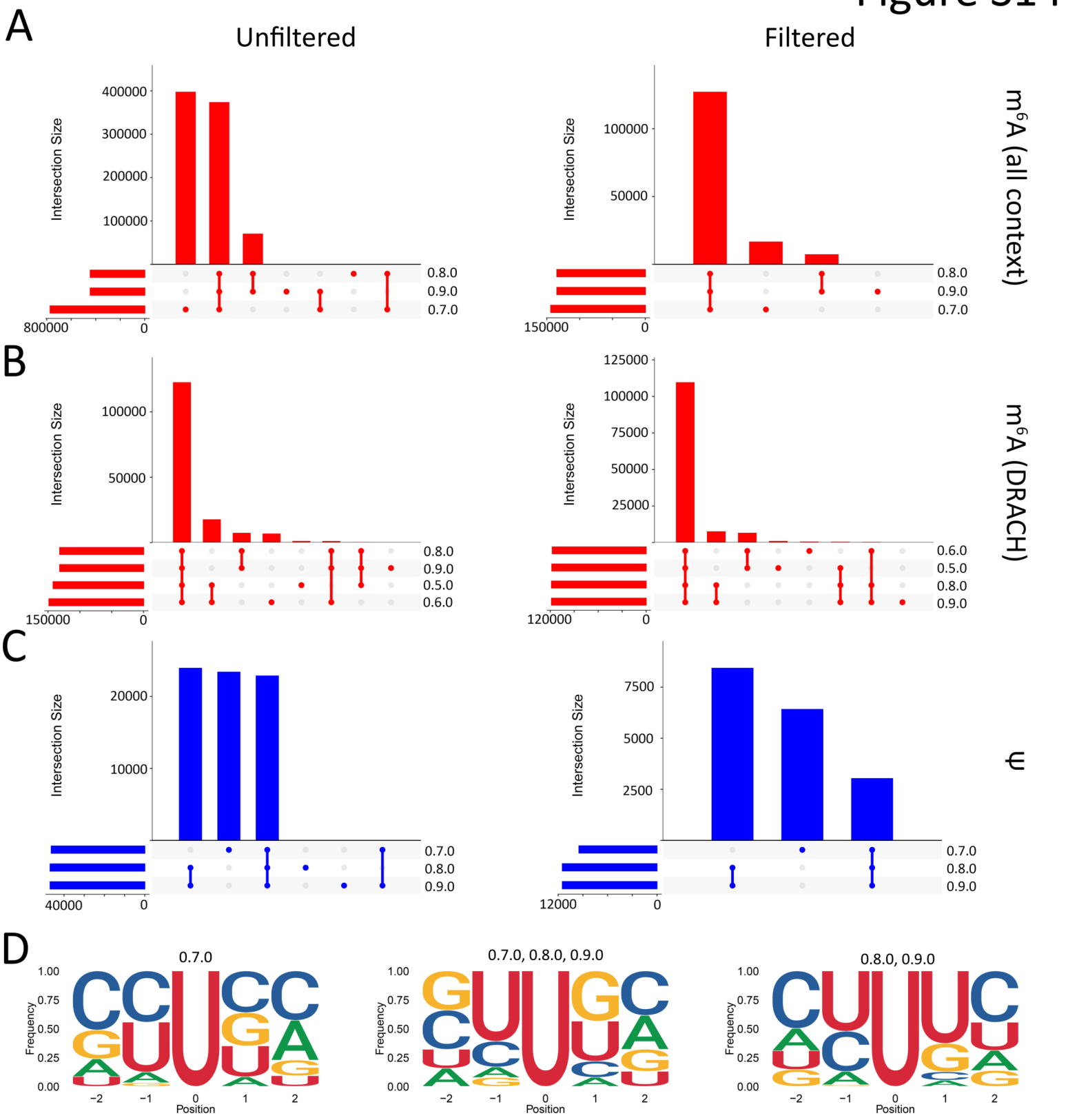

A
Unfiltered
Filtered
m6A (all context)
Intersection Size
400000
300000
200000
100000
800000
0
0.8.0
0.9.0
0.7.0
100000
50000
150000
B
m6A (DRACH)
125000
75000
25000
120000
0.5.0
0.6.0
C
ψ
20000
10000
40000
7500
5000
2500
12000
D
0.7.0
0.7.0, 0.8.0, 0.9.0
0.8.0, 0.9.0
Frequency
1.00
0.75
0.50
0.25
0.00
−2
−1
0
1
2
Position

Figure S15

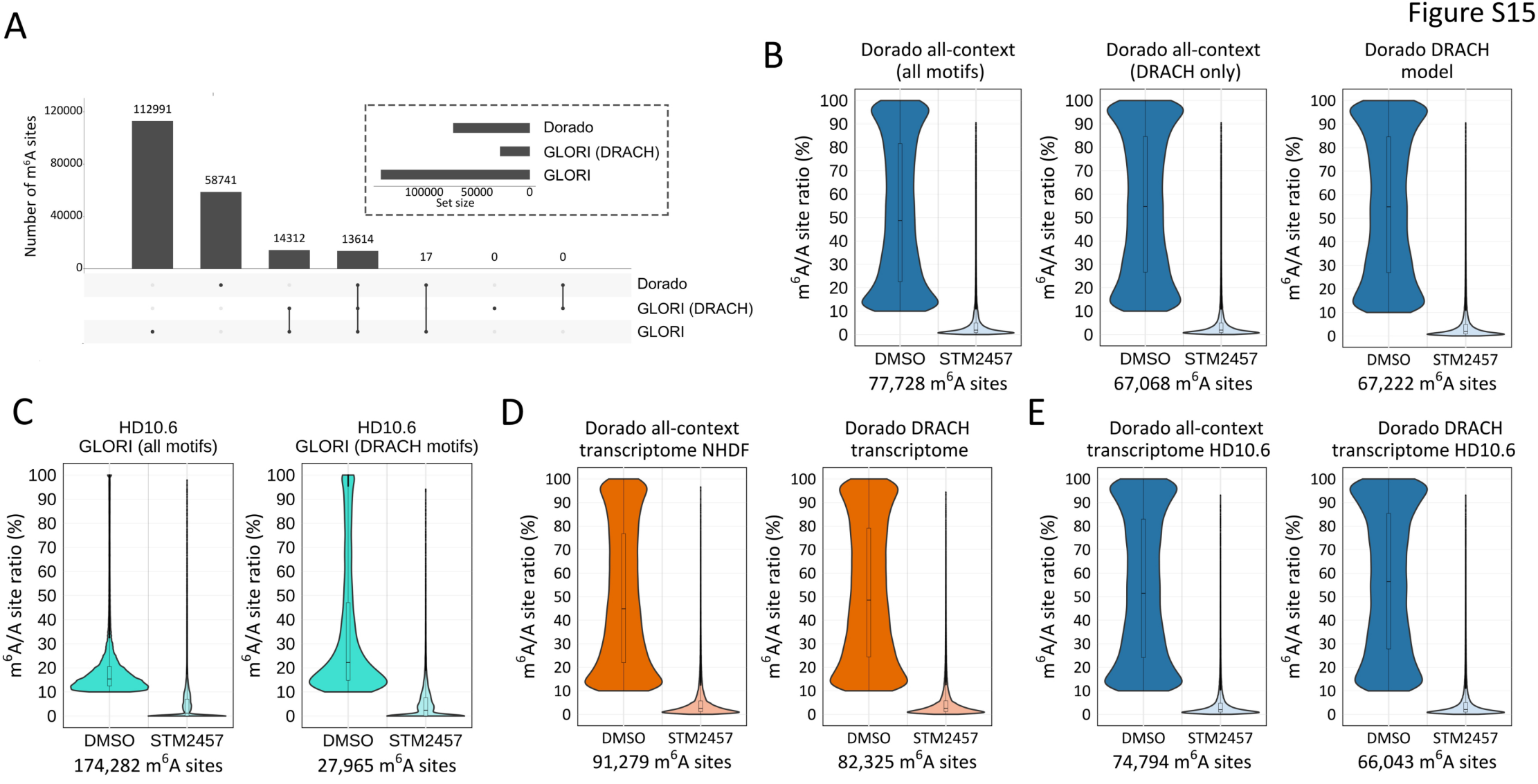
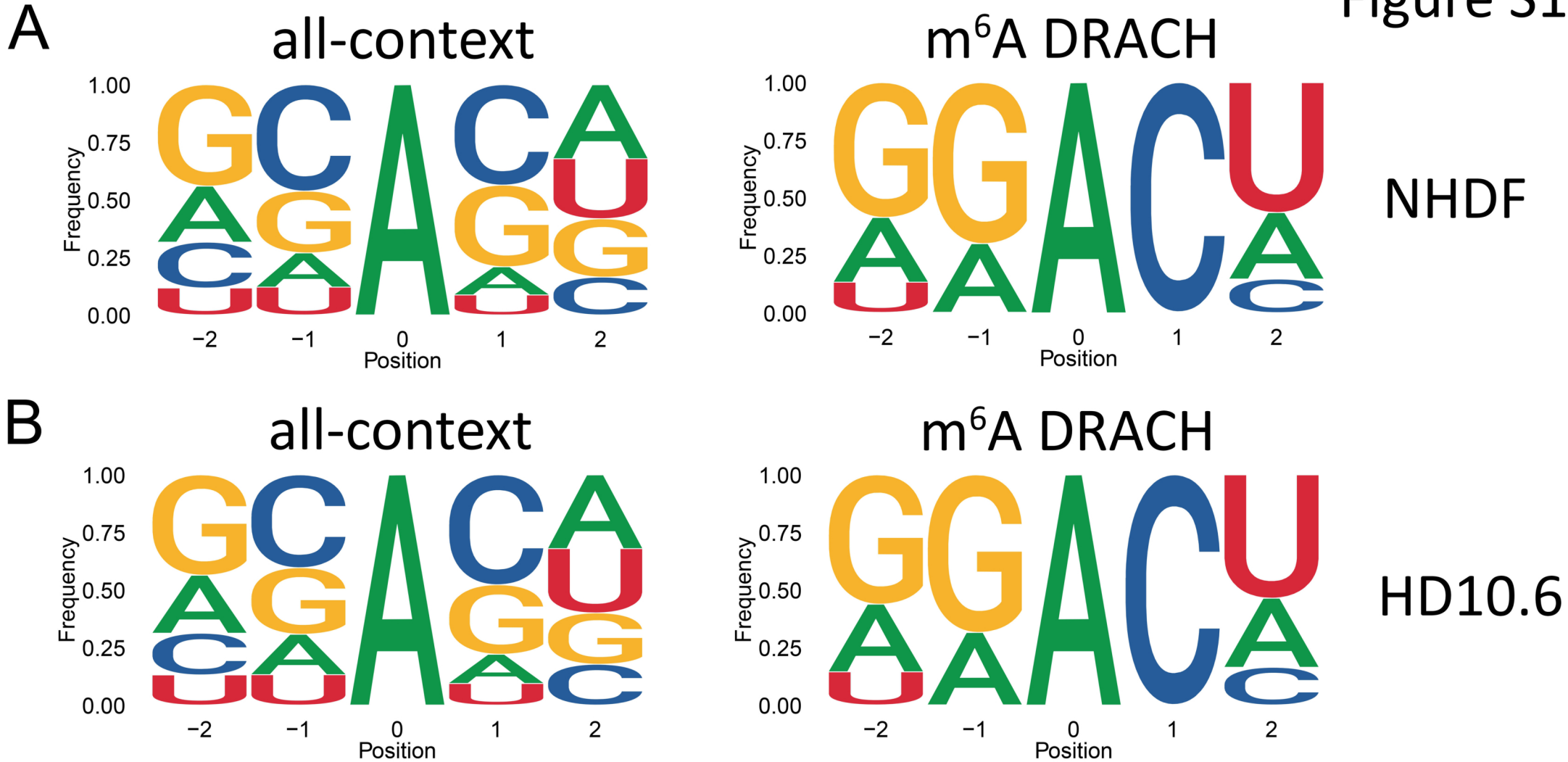

A
Number of m6A sites
120000
80000
40000
0
112991
58741
14312
13614
17
0
0
Dorado
GLORI (DRACH)
GLORI
100000
50000
0
Set size
B
Dorado all-context (all motifs)
Dorado all-context (DRACH only)
Dorado DRACH model
m6A/A site ratio (%)
DMSO
STM2457
77,728 m6A sites
67,068 m6A sites
67,222 m6A sites
C
HD10.6 GLORI (all motifs)
HD10.6 GLORI (DRACH motifs)
174,282 m6A sites
27,965 m6A sites
D
Dorado all-context transcriptome NHDF
Dorado DRACH transcriptome
91,279 m6A sites
82,325 m6A sites
E
Dorado all-context transcriptome HD10.6
Dorado DRACH transcriptome HD10.6
74,794 m6A sites
66,043 m6A sites

Figure S16


A
all-context
m6A DRACH
NHDF
B
all-context
m6A DRACH
HD10.6
Frequency
1.00
0.75
0.50
0.25
0.00
−2
−1
0
1
2
Position

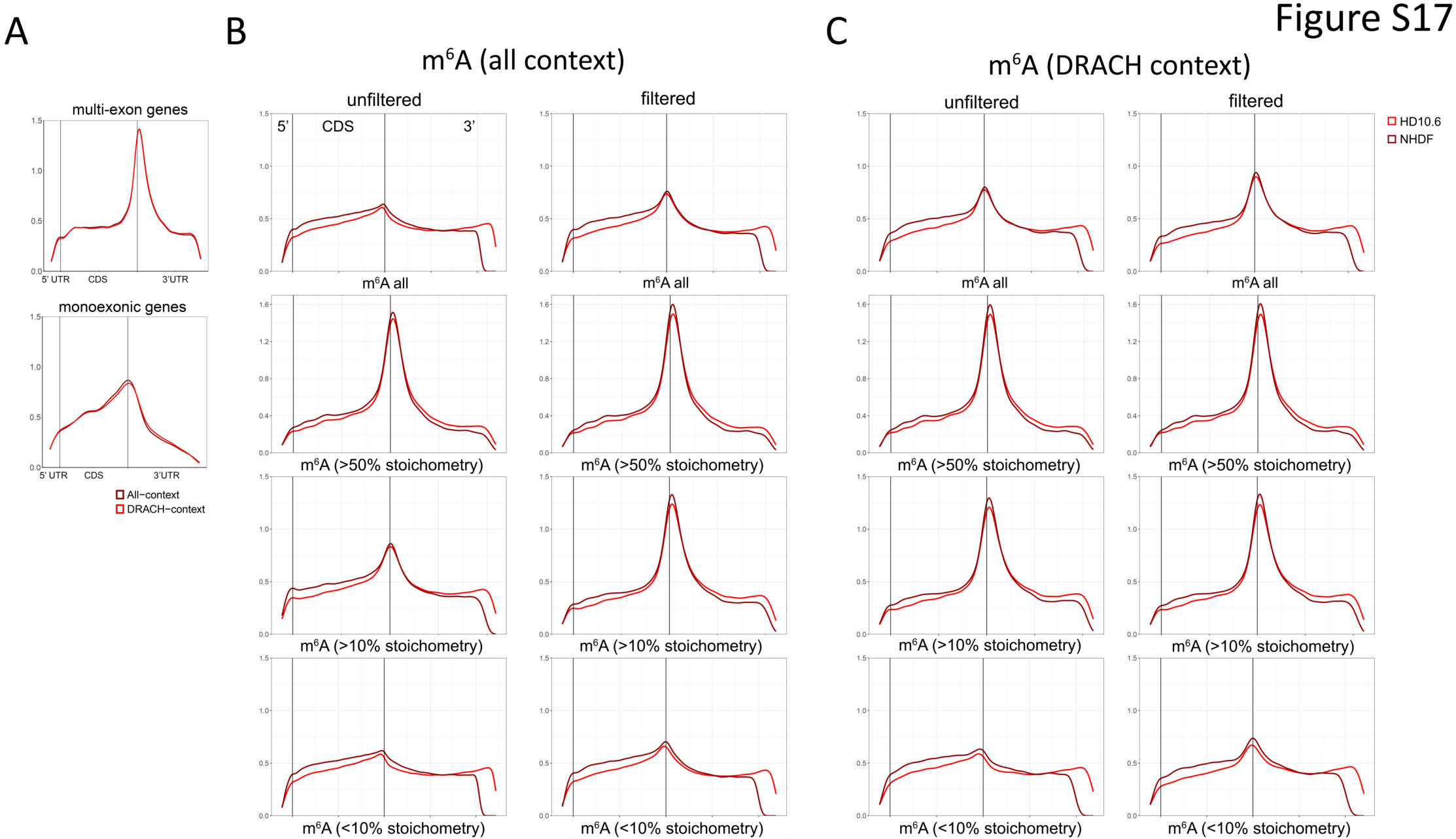


Figure S17
A
multi-exon genes
5' UTR
CDS
3'UTR
monoexonic genes
All−context
DRACH−context
B
m6A (all context)
unfiltered
filtered
5'
CDS
3'
m6A all
m6A (>50% stoichometry)
m6A (>10% stoichometry)
m6A (<10% stoichometry)
C
m6A (DRACH context)
HD10.6
NHDF