

Mapping of RNA modifications by direct Nanopore sequencing and JACUSA2

Christoph Dieterich^{*1,2,3}, Amina Lemsara^{1,2}, and Isabel Naarmann-de Vries^{1,2,3}

¹Klaus Tschira Institute for Integrative Computational Cardiology, University Heidelberg, 69120 Heidelberg, Germany

²Department of Internal Medicine III (Cardiology, Angiology, and Pneumology), University Hospital Heidelberg, 69120 Heidelberg, Germany

³German Centre for Cardiovascular Research (DZHK)-Partner Site Heidelberg/Mannheim, 69120 Heidelberg, Germany

Abstract

to be written

Keywords: Bayesian, 10X Genomics, Cell barcode assignment, Nonsense-mediated mRNA decay (NMD)

INTRODUCTION

Chemical modifications on DNA and histones, also known as epigenetics marks, strongly impact gene expression during cell differentiation and in several other biological programs. In the 1970s, it was recognized that RNA is also subjected to extensive covalent modification, and studies in the late 1980s revealed the widespread deamination of bases (termed RNA editing), which can lead to recoding if it occurs within coding sequences. Impressive development in the RNA modification field occurred during the past eight years, with the discovery of an extensive layer of base modifications in mRNAs. These can influence gene expression and have been already shown to be involved in primary cellular programs such as stem cell differentiation, response to stress, and the circadian clock. The study of RNA modifications and their effects is now referred to as epitranscriptomics, and it reveals striking similarities to what is known for epigenomics. To date thirteen distinct modifications have been identified on mRNA transcripts [Anreiter et al., 2021]. These modifications are catalyzed by a variety of dedicated enzymes and can be divided into two classes: modifications of cap-adjacent nucleotides and internal modifications.

^{*}christoph.dieterich@uni-heidelberg.de

32 In contrast to the m7G cap, the impact of internal modifications on gene
 33 regulation has been less studied apart from RNA editing, which is mediated
 34 by RNA deaminases (e.g. the ADAR family). The most widespread in-
 35 ternal mRNA modification is N6-methyladenosine (m6A). By modulating
 36 the processing of mRNA, m6A can regulate a wide range of physiological
 37 processes and its alteration has been linked to several diseases Roignant
 38 and Soller [2017], Zaccara et al. [2019], Shi et al. [2019]. The modification is
 39 catalyzed co-transcriptionally by a Mega-Dalton methyltransferase complex,
 40 which includes the heterodimer METTL3-METTL14 and other associated
 41 subunits Garcias Morales and Reyes [2021]. This modification is reversible
 42 since two proteins of the AlkB-family demethylases can remove m6A from
 43 mRNA transcripts [Jia et al., 2011, Zheng et al., 2013]. In mammals, m6A
 44 preferentially localizes within long internal exons and at the beginning of
 45 terminal exons at so-called DRACH motif (D = A/G/U, R = A/G, H =
 46 A/C/U) sites [Dominissini et al., 2012, Meyer et al., 2012, Ke et al., 2015].
 47 Once deposited, m6A is recognized by several reader proteins that can af-
 48 fect the fate of mRNA transcripts in nearly every step of the mRNA life
 49 cycle, which includes alternative splicing [Adhikari et al., 2016, Roundtree
 50 et al., 2017]. The best-described readers are the YTH domain family of
 51 proteins that decode the signal and mediate m6A functions. By affecting
 52 RNA structure, m6A can also indirectly influence the association of addi-
 53 tional RNA-binding proteins (RBPs) and the assembly of larger messenger
 54 ribonucleoprotein particles (mRNPs).

55 Several approaches have been presented to map RNA modifications on
 56 RNA. Herein, we focus on mRNA modification site detection in general and
 57 on m6A in particular where antibody-based protocols (miCLIP), methylation-
 58 sensitive restriction enzyme assays (MazF) or transgenic approaches (TRIBE,
 59 DART) have been presented. All of the aforementioned approaches rely on
 60 high-throughput sequencing on the Illumina platform. This typically in-
 61 volves cDNA synthesis by reverse transcription and PCR-based library am-
 62 plification. One recent addition to the tool is direct RNA single molecule
 63 sequencing on the Oxford Nanopore Technology platform. While or software
 64 workflow is able to deal with Illumina and Nanopore-based approaches, the
 65 latter is the principal topic of our methods article.

66 MATERIALS

67 ONT direct RNA sequencing

- 68 1. 500 ng polyA⁺ RNA isolated from total RNA e.g. with Oligotex
 69 mRNA kit (Qiagen) or Dynabeads oligo dT₂₅ beads (Thermo Fisher
 70 Scientific) or *in vitro* transcriptome sample. Store RNA at -80 °C and
 71 the mRNA purification kit as recommended by the manufacturer.

- 72 2. Nuclease-free water. Store at room temperature.
- 73 3. Direct RNA-sequencing kit (SQK-RNA002, Oxford Nanopore Tech-
74 nologies). Store at -20 °C.
- 75 4. NEBNext Quick Ligation Reaction Buffer (New England Biolabs).
76 Store at -20 °C.
- 77 5. T4 DNA Ligase (New England Biolabs). Store at -20 °C.
- 78 6. dNTP Mix (10 mM each). Store at -20 °C.
- 79 7. SuperScript IV Reverse Transcriptase (Thermo Fisher Scientific). Store
80 at -20 °C.
- 81 8. Agencourt RNAClean XP beads (Beckman Coulter). Store at 4 °C.
- 82 9. 70 % ethanol, freshly prepared.
- 83 10. Qubit dsDNA HS assay kit and Qubit Fluorometer (Thermo Fisher
84 Scientific).
- 85 11. Flow cell priming kit (EXP-FLP002, Oxford Nanopore Technologies).
86 Store at -20 °C.
- 87 12. Thermocycler.
- 88 13. Gentle rotator mixer.
- 89 14. Magnetic stand for 1.5 ml tubes.
- 90 15. 1.5 ml DNA LoBind tubes (Eppendorf), 0.2 ml PCR tubes.
- 91 16. MinION or GridION sequencing device and MinION R9.4.1 Flow cells
92 (FLO-MIN106D, Oxford Nanopore Technologies). Store Flow cells at
93 4 °C.

94 **Preparation of an *in vitro* transcriptome sample**

- 95 1. 100 ng polyA⁺ RNA isolated from total RNA e.g. with Oligotex
96 mRNA kit (Qiagen) or Dynabeads oligo dT₂₅ beads (Thermo Fisher
97 Scientific). Store RNA at -80 °C and the mRNA purification kit as
98 recommended by the manufacturer
- 99 2. 10 μ M oligo(dT)-VN RT primer. TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN.
100 Store at -20 °C.
- 101 3. 20 μ M template switching oligo (TSO). ACTCTAATACGACTCAC-
102 TATAGGGAGAGGGCrGrG+G. Store at -20 °C.

- 103 4. 10 μ M T7 extension primer. GCTCTAATACGACTCACTATAGG.
104 Store at -20 °C.
- 105 5. Nuclease-free water. Store at room temperature.
- 106 6. dNTP Mix (10 mM each). Store at -20 °C.
- 107 7. Template Switching RT Enzyme Mix (New England Biolabs). Store
108 at -20 °C.
- 109 8. Q5 Hot Start High-Fidelity 2X Master Mix (New England Biolabs).
110 Store at -20 °C.
- 111 9. RNase H (5,000 U/ml) (New England Biolabs). Store at -20 °C.
- 112 10. NucleoSpin Gel and PCR Clean-up, Mini kit for gel extraction and
113 PCR clean up (Macherey-Nagel) or equivalent. Store at room temper-
114 ature.
- 115 11. MEGAscript T7 transcription kit (Thermo Fisher Scientific). Store at
116 -20 °C.
- 117 12. RNA Clean & Concentrator-25 kit (Zymo Research). Store at room
118 temperature.
- 119 13. Thermocycler.
- 120 14. Table top centrifuge for 1.5 ml tubes.
- 121 15. Nanodrop spectrophotometer or equivalent.
- 122 16. 0.2 ml PCR tubes, 1.5 ml DNA LoBind tubes (Eppendorf).

123 **Hardware requirements**

124 All analyses have been performed/tested on two alternative hardware sys-
125 tems: a standard Linux desktop computer or an Apple iMac (Retina 5K,
126 ultimo 2014). The workflow requires a multi-core processor system with
127 minimal main memory of 16GB RAM and several GBs of free disk space
128 (depending on data set size).

129 **Software dependencies and installation**

130 Our analysis workflow has few requirements, which are detailed in Table 2.
131 Specifically, to execute our workflow, the following prerequisites are neces-
132 sary: a BASH shell, a JAVA runtime environment, a working PERL and
133 R installation. Additional i.e. non-standard software to process and map
134 Nanopore reads (bedtools, samtools and Minimap2) are obligatory, while

135 the installation of a Nanopore read simulator (NanoSim) is optional and de-
136 pends on your use case. Table ?? lists some additional R packages, which are
137 required to run the R code. Detailed instructions on how to setup are found
138 under https://github.com/dieterich-lab/MiMB_JACUSA2_chapter

139 METHODS

140 Overview Figure 1

141 Nanopore direct RNA sequencing

- 142 1. Adjust 500 ng polyA⁺ RNA to a total volume of 9 μ l with nuclease-
143 free water. Complete RT adapter ligation reaction (in 0.2 ml PCR
144 tube) with 3 μ l NEBNext Quick Ligation Reaction Buffer, 0.5 μ l
145 RNA CS (RCS, from SQK-RNA002), 1 μ l RT-Adapter (RTA, from
146 SQK-RNA002) and 1.5 μ l T4 DNA Ligase. Incubate 10 min at room
147 temperature.
- 148 2. Prepare reverse transcription master mix on ice during ligation: 9 μ l
149 nuclease-free water, 2 μ l 10 mM dNTPs, 8 μ l 5x SuperScript IV first
150 strand buffer, 4 μ l 0.1 mM DTT.
- 151 3. Add the reverse transcription master mix to the ligation reaction and
152 mix by pipetting. Add 2 μ l SuperScript IV reverse transcriptase and
153 mix by pipetting. Incubate in a thermocycler with the following pro-
154 tocol: 50 min at 50 °C, 10 min at 70 °C, cool down to 4 °C.
- 155 4. Let the Agencourt RNAClean XP beads come to room temperature
156 during reverse transcription. Carefully resuspend beads before use.
157 Transfer reaction to a 1.5 ml DNA LoBind tube and mix with 72 μ l
158 Agencourt RNAClean XP beads. Incubate 5 min at room temperature
159 on a gentle rotator mixer.
- 160 5. Collect beads on a magnetic stand and remove supernatant. Wash
161 pelleted beads two times (30 sec) with 200 μ l freshly prepared 70 %
162 ethanol. Remove supernatant. Spin sample down and place on magnet
163 again. Remove any residual ethanol.
- 164 6. Resuspend beads in 20 μ l nuclease-free water by gentle flicking and
165 incubate 5 min at room temperature on a gentle rotator mixer. Collect
166 beads on a magnetic stand and transfer 20 μ l eluate in a fresh 1.5 ml
167 DNA LoBind tube.
- 168 7. For ligation of the RMX adapter, add the following to 20 μ l eluate: 8
169 μ l NEBNext Quick Ligation Reaction Buffer, 6 μ l RMX (from SQK-
170 RNA002), 3 μ l nuclease-free water, 3 μ l T4 DNA Ligase. Mix by
171 pipetting and incubate 10 min at room temperature.

- 172 8. Add 40 μ l carefully resuspended Agencourt RNAClean XP beads to
173 the reaction and mix by pipetting. Incubate 5 min at room tempera-
174 ture on a gentle rotator mixer.
- 175 9. Collect beads on a magnetic stand and remove supernatant. Wash
176 pelleted beads two times with 150 μ l wash buffer (WSB, from SQK-
177 RNA002). Resuspend beads by flicking, spin down and return to mag-
178 netic stand. Remove supernatant from pelleted beads.
- 179 10. Resuspend beads in 21 μ l elution buffer (EB, from SQK-RNA002) by
180 gentle flicking and incubate 5 min at room temperature on a gentle
181 rotator mixer. Pellet beads on a magnetic stand and transfer 21 μ l
182 eluate in a fresh 1.5 ml DNA LoBind tube.
- 183 11. Quantify 1 μ l of the library on a Qubit fluorometer with the Qubit
184 dsDNA HS kit according to the manufacturerers protocol. Concentra-
185 tion should be usually in the range of 5 - 10 ng/ μ l.
- 186 12. Insert MinION R9.4.1 Flow cell in the MinION or GridION sequenc-
187 ing device and perform Flow cell check in the MinKNOW software.
188 For successful sequencing of mammalian polyA⁺ RNA at least 1,000
189 available pores are recommended.
- 190 13. Prepare Priming Mix by adding 30 μ l flush tether (FLT, from EXP-
191 FLP002) to a vial of flush buffer (FB, from EXP-FLP002) and mix by
192 pipetting. Open priming port. Remove air bubble from priming port
193 by inserting the tip of a P1000 pipette into the priming port and slowly
194 dialing up, until a small volume of storage buffer enters the pipette
195 tip. Load 800 μ l Priming Mix via the priming port and carefully avoid
196 introduction of air bubbles. Close the priming port and wait for 5 min.
- 197 14. Mix 20 μ l library with 17.5 μ l nuclease-free water and 37.5 μ l RNA run-
198 ning buffer (RRB, from SQK-RNA002) and mix by pipetting. Open
199 the priming port and the sample port. Load 200 μ l Priming Mix via
200 the priming port. Mix library by pipetting just before loading and
201 load dropwise via the sample port. Carefully avoid introduction of air
202 bubbles. Close the sample port and the priming port.
- 203 15. Start sequencing for 48 to 72 h in the MinKNOW software. Choose
204 direct RNA-sequencing kit and high-accuracy basecalling as paramet-
205 ers. We recommend to adjust the output filter to a minimum Q score
206 of 7 (instead of 9).

207 Preparation of an *in vitro* transcriptome sample

208 The *in vitro* transcriptome sample is prepared based on a protocol published
209 by Zhang *et al.* Zhang et al. [2021] with some modifications.

- 210 1. Adjust 100 ng polyA⁺ RNA to a total volume of 6 μ l with nuclease-
211 free water. Add 1 μ l each of 10 μ M oligo(dT)-VN RT primer and 10
212 mM dNTPs. Mix by pipetting and incubate in a thermocycler: 5 min
213 at 75 °C, 2 min at 42 °C, cool to 4 °C.
- 214 2. Assemble 2.5 μ l 4x template switching RT buffer, 0.5 μ l 20 μ M TSO,
215 1 μ l 10x template switching RT enzyme mix and mix by pipetting.
216 Combine with 6 μ l RNA and incubate in a thermocycler: 90 min at
217 42 °C, 10 min at 68 °C, cool to 4 °C.
- 218 3. For Second strand synthesis add to First strand synthesis reaction: 50
219 μ l Q5 Hot Start High-Fidelity 2X Master Mix, 5 μ l RNase H, 2 μ l 10
220 μ M T7 extension primer, 33 μ l nuclease-free water. Mix by pipetting
221 and incubate in a thermocycler: 15 min at 37 °C, 1 min at 95 °C, 10
222 min at 65 °C, cool to 4 °C.
- 223 4. Purify double stranded cDNA with NucleoSpin Gel and PCR Clean-up
224 kit according to the manufacturer's protocol and elute in 20 μ l elution
225 buffer. Determine concentration on a Nanodrop spectrophotometer.
226 cDNA may be stored at -20 °C.
- 227 5. Combine 8 μ l cDNA for *in vitro* transcription with 2 μ l each of ATP,
228 GTP, CTP, UTP, 10x reaction buffer and enzyme mix from the MEGAscript
229 T7 transcription kit. Incubate 3 h at 37 °C.
- 230 6. Digest template DNA by addition of 1 μ l Turbo DNase. Mix by pipet-
231 ting and incubate 15 min at 37 °C.
- 232 7. Adjust reaction volume to 100 μ l with nuclease-free water and clean up
233 with RNA Clean & Concentrator-25 kit according to the manufactur-
234 ers protocol, using two volumes of adjusted RNA binding buffer (1:1
235 RNA binding buffer : ethanol). Elute RNA in 25 μ l nuclease-free wa-
236 ter. Determine RNA concentration on a Nanodrop spectrophotometer.
237 Store at -80 °C.

238 Nanopore read processing

- 239 1. JACUSA2 requires indexed BAM files. In order to create a BAM file
240 index for an existing file align.bam, use samtools.
- 241

```
$ samtools index align.bam
```
- 242 2. Add the MD tag field to the BAM files. This requires the reference
243 sequence reference.fasta. The MD tag field stores information on mis-
244 matched and deleted reference bases.
- 245

```
$ samtools calmd -b align.bam reference.fasta > align_md.bam
```

246 Use Case 1: Comparison of wildtype and knock-out samples

247 The benchmark is composed of two sample types: wildtype and knock-out
248 samples for Hek293 cell line with two replicates. Given the preprocessed
249 mapped reads as input (BAM files) 'HEK293T-WT-rep2.bam, HEK293T-
250 WT-rep3.bam, HEK293T-KO-rep2.bam,HEK293T-KO-rep3.bam',

- 251 1. Identify read error profile: run JACUSA2 with paired samples option
252 (call-2). One should specify the corresponding library information and
253 may filter reads according to many parameters. Here, we are consider-
254 ing all positions with read coverage ≥ 4 . The output resume read error
255 profile where the format is a combination of BED6 with JACUSA2
256 methods specific columns and common info columns: "info", "filter",
257 and "ref". See JACUSA2 manual for more details on JACUSA2 filter
258 and output options.

```
259 JACUSA2 2.0.0-RC22 call-2 -m 1 -q 1 -c 4 -p 10 -D -I -a D,Y -P1 FR-SECONDSTRAND  
260 -P2 FR-SECONDSTRAND -r WT_vs_KO_2samp_RC22_call2_result.out HEK293T-WT-rep2.bam  
261 HEK293T-KO-rep2.bam,HEK293T-KO-rep3.bam
```

- 262 2. Preprocess JACUSA2 output: given the JACUSA2 output, select non-
263 overlapping sites of homo-polymer regions (JACUSA filter: Y) and
264 within a 5mer of a central A nucleotide flanked by 2 adjacent random
265 nucleotides (NNANN). Each position is represented by the insertion,
266 deletion and mismatch scores and the position number within the 5mer
267 specific context. The README_processing.sh bash script perform the
268 preprocessing and produces a txt file call2_SitesExt2_indel_slim2.txt
269 resuming the features for selected positions and a separate file rep-
270 resenting the 5mer bases named checkMotif_reformat.txt. One may
271 precise the path to outputs within the command.

```
272 bash README_processing.sh WT_vs_KO_RC22_call2_result.out hg38.genome GRCh38_96.
```

- 273 3. Extract 5mer features: extract features representing the insertion,
274 deletion and mismatch within the specific 5mer context. To do so,
275 run the R script HEK293_data_prep.R. This will produce an R object
276 named BigTable.rds, representing the matrix of positions x 15 features
277 which correspond to the insertion, deletion and mismatch scores for
278 the observed position and its two flanking positions. To run the script,
279 precise the path to outputs that contains already the reformatted data
280 and provide also the sample's name as a label of the analysis.

```
281 Rscript Code/HEK293_data_prep.R path_to_output WT_vs_KO_RC22_call2_result.out
```


- 282 4. Extract m6A modification patterns: now that one got the matrix of
283 Sites X Features, the next step is to extract patterns allowing to predict
284 m6A modified positions. To this end, the non-negative matrix factor-
285 ization (NMF) analysis is suggested. The R script HEK293_data_prep_step2.R
286 allows to generate patterns from a subset of data associated to previ-
287 ously reported m6A positions. Here, the unsupervised pattern learn-
288 ing approach is based on 2401 m6A sites. Based on the silhouette and
289 cophenetic correlation indices, we could identify an optimal factoriza-
290 tion rank of 7 (fig. 4).
- 291 5. Predict m6A modifications: The cumulative density function of pat-
292 tern 3 scores can be used to predict novel m6A sites.

293 Use Case 2: Comparison of wildtype and IVT samples

294 Christoph

295 Use Case 3: Comparison of wildtype to simulated IVT sample

296 Christoph

297 NOTES

298 Tips and Tricks

299 ACKNOWLEDGMENTS

300 The authors would like to thank Etienne Boileau, Thiago Britto Borges,
301 Tobias Jakobi for proof-reading and comments. The authors are grateful
302 to Marek Franitza for running the experiments on the 10x platform and to
303 Christian Becker for running ONT sequencing. This work was supported by
304 Informatics for Life funded by the Klaus Tschira Foundation.

305 REFERENCES

- 306 Samir Adhikari, Wen Xiao, Yong-Liang Zhao, and Yun-Gui Yang. m(6)a:
307 Signaling for mrna splicing. *RNA biology*, 13:756–759, September 2016.
308 ISSN 1555-8584. doi: 10.1080/15476286.2016.1201628.
- 309 Ina Anreiter, Quoseena Mir, Jared T. Simpson, Sarath C. Janga, and
310 Matthias Soller. New twists in detecting mrna modification dynamics.
311 *Trends in biotechnology*, 39:72–89, January 2021. ISSN 1879-3096. doi:
312 10.1016/j.tibtech.2020.06.002.

313 Dan Dominissini, Sharon Moshitch-Moshkovitz, Schraga Schwartz, Mali
 314 Salmon-Divon, Lior Ungar, Sivan Osenberg, Karen Cesarkas, Jasmine
 315 Jacob-Hirsch, Ninette Amariglio, Martin Kupiec, Rotem Sorek, and
 316 Gideon Rechavi. Topology of the human and mouse m6a rna methylomes
 317 revealed by m6a-seq. *Nature*, 485:201–206, April 2012. ISSN 1476-4687.
 318 doi: 10.1038/nature11112.

319 David Garcias Morales and José L. Reyes. A birds’-eye view of the activ-
 320 ity and specificity of the mrna m, javax.xml.bind.jaxbelement@6d66739e,
 321 a methyltransferase complex. *Wiley interdisciplinary reviews. RNA*, 12:
 322 e1618, January 2021. ISSN 1757-7012. doi: 10.1002/wrna.1618.

323 Guifang Jia, Ye Fu, Xu Zhao, Qing Dai, Guanqun Zheng, Ying Yang,
 324 Chengqi Yi, Tomas Lindahl, Tao Pan, Yun-Gui Yang, and Chuan He.
 325 N6-methyladenosine in nuclear rna is a major substrate of the obesity-
 326 associated fto. *Nature chemical biology*, 7:885–887, October 2011. ISSN
 327 1552-4469. doi: 10.1038/nchembio.687.

328 Shengdong Ke, Endalkachew A. Alemu, Claudia Mertens, Emily Conn Gant-
 329 man, John J. Fak, Aldo Mele, Bhagwattie Haripal, Ilana Zucker-Scharff,
 330 Michael J. Moore, Christopher Y. Park, Cathrine Broberg Vågbø, Anna
 331 Kusnierz, Arne Klungland, James E. Darnell, and Robert B. Darnell.
 332 A majority of m6a residues are in the last exons, allowing the potential
 333 for 3’ utr regulation. *Genes & development*, 29:2037–2053, October 2015.
 334 ISSN 1549-5477. doi: 10.1101/gad.269415.115.

335 Kate D. Meyer, Yogesh Saletore, Paul Zumbo, Olivier Elemento, Christo-
 336 pher E. Mason, and Samie R. Jaffrey. Comprehensive analysis of mrna
 337 methylation reveals enrichment in 3’ utrs and near stop codons. *Cell*, 149:
 338 1635–1646, June 2012. ISSN 1097-4172. doi: 10.1016/j.cell.2012.05.003.

339 Jean-Yves Roignant and Matthias Soller. m,
 340 javax.xml.bind.jaxbelement@8cec19d, a in mrna: An ancient mech-
 341 anism for fine-tuning gene expression. *Trends in genetics : TIG*, 33:
 342 380–390, June 2017. ISSN 0168-9525. doi: 10.1016/j.tig.2017.04.003.

343 Ian A. Roundtree, Molly E. Evans, Tao Pan, and Chuan He. Dynamic rna
 344 modifications in gene expression regulation. *Cell*, 169:1187–1200, June
 345 2017. ISSN 1097-4172. doi: 10.1016/j.cell.2017.05.045.

346 Hailing Shi, Jiangbo Wei, and Chuan He. Where, when, and how:
 347 Context-dependent functions of rna methylation writers, readers, and
 348 erasers. *Molecular cell*, 74:640–650, May 2019. ISSN 1097-4164. doi:
 349 10.1016/j.molcel.2019.04.025.

350 Sara Zaccara, Ryan J. Ries, and Samie R. Jaffrey. Reading, writing and
351 erasing mrna methylation. *Nature reviews. Molecular cell biology*, 20:608–
352 624, October 2019. ISSN 1471-0080. doi: 10.1038/s41580-019-0168-5.

353 Zhang Zhang, Tao Chen, Hong-Xuan Chen, Ying-Yuan Xie, Li-Qian Chen,
354 Yu-Li Zhao, Biao-Di Liu, Lingmei Jin, Wutong Zhang, Chang Liu,
355 et al. Systematic calibration of epitranscriptomic maps using a synthetic
356 modification-free rna library. *Nature Methods*, 18(10):1213–1222, 2021.

357 Guanqun Zheng, John Arne Dahl, Yamei Niu, Peter Fedorcsak, Chun-Min
358 Huang, Charles J. Li, Cathrine B. Vågbo, Yue Shi, Wen-Ling Wang, Shu-
359 Hui Song, Zhike Lu, Ralph P. G. Bosmans, Qing Dai, Ya-Juan Hao, Xin
360 Yang, Wen-Ming Zhao, Wei-Min Tong, Xiu-Jie Wang, Florian Bogdan,
361 Kari Furu, Ye Fu, Guifang Jia, Xu Zhao, Jun Liu, Hans E. Krokan, Arne
362 Klungland, Yun-Gui Yang, and Chuan He. Alkbh5 is a mammalian rna
363 demethylase that impacts rna metabolism and mouse fertility. *Molecular*
364 *cell*, 49:18–29, January 2013. ISSN 1097-4164. doi: 10.1016/j.molcel.2012.
365 10.015.

FIGURE CAPTIONS

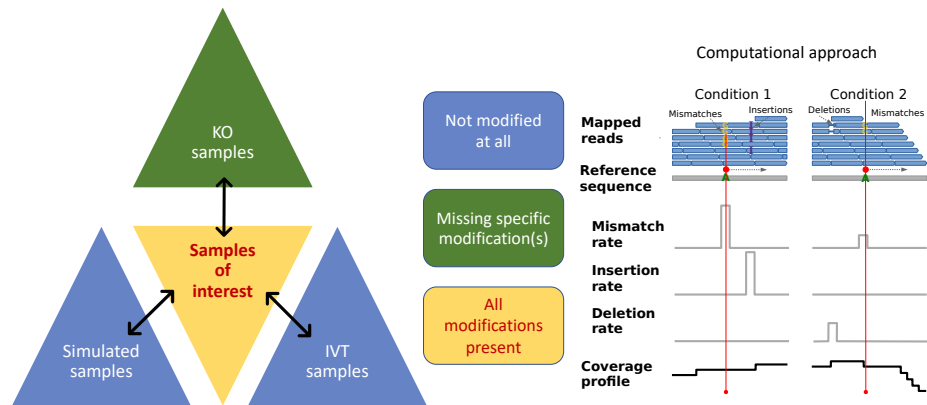


Figure 1: **General outline of RNA modification detection by JACUSA2.** A key feature of our approach is that multiple replicates can be compared as shown on the left. Samples of interests where all modifications are present could be compared with either KO samples where the modification of interest is missing or IVT/simulated samples where all modifications are absent. Read stacks (in blue) are compared head-to-head as shown on the right.

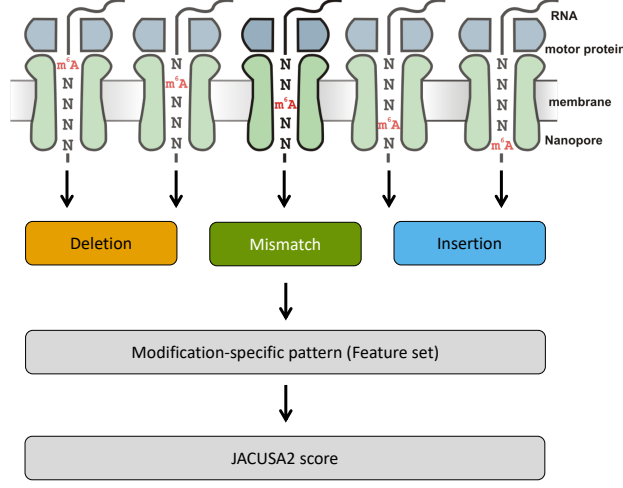


Figure 2: **Motivation of 5mer context for RNA modification mapping.** The nanopore covers 5 consecutive RNA residues. That is why we consider a 5mer context and derive 3 principal features for every position within a given 5 mer (15 features in total, with a central A residue in this example). We evaluate each feature set by previously learned patterns and compute a final score for modification site detection.

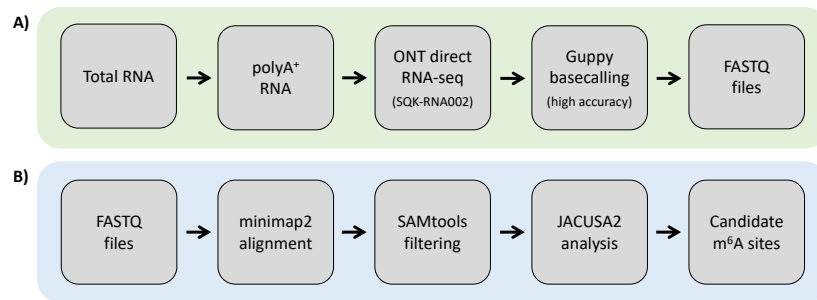


Figure 3: **Experimental and computational workflow.** tbd

FIGURE CAPTIONS

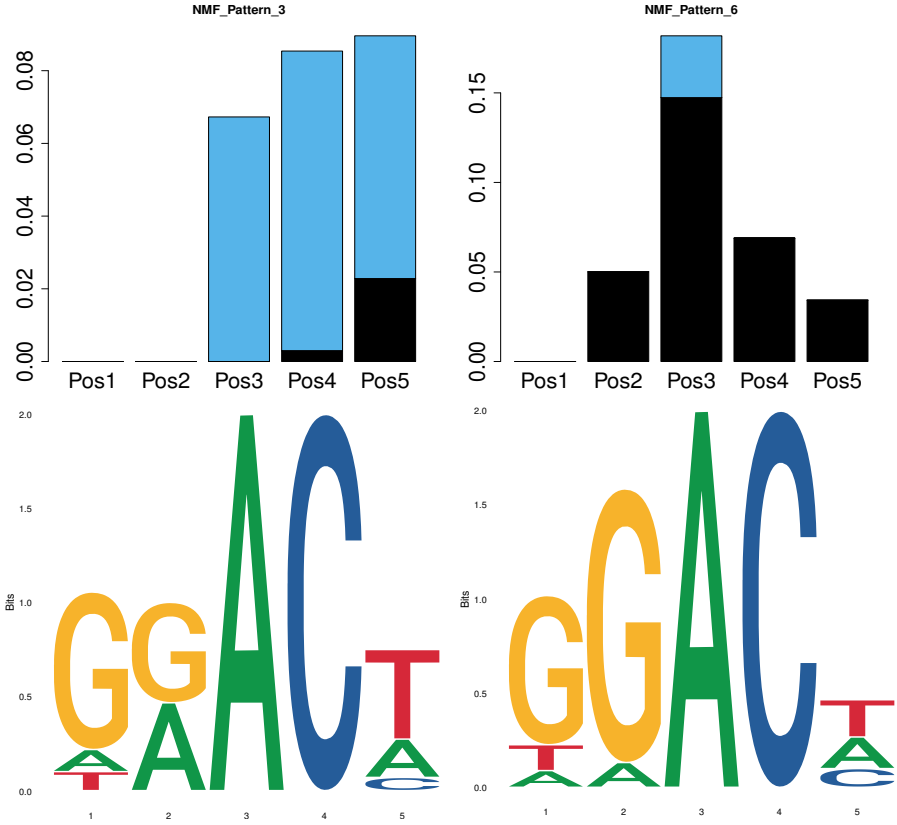


Figure 4: Main patterns for WT vs KO case.

Software	Version	Description
Minimap2	https://github.com/lh3/minimap2 v2.22 or later	https://lh3.github.io/minimap2/
samtools	https://github.com/samtools/samtools v1.12 or later	http://samtools.github.io/
JAVA	openjdk 11.0.12 2021-07-20 - JAVA 11 or later	OpenJDK Runtime Environment
R	https://www.r-project.org/ version 3.5.1 or later	The R Project for Statistical Computing
PERL	https://www.perl.org/ version 5.28.1 or later	Perl is a highly capable, feature-rich programming language
BASH, sed, awk	should be part of your Linux distribution	Misc.
bedtools	https://github.com/arq5x/bedtools2 version 2.29.2 or later	Perl is a highly capable, feature-rich programming language
NanoSim	https://github.com/bcgsc/NanoSim version 3.0.2 or later (optional)	NanoSim is a fast and scalable read simulator that captures the technology-specific features of ONT data

Table 1: **Software dependencies** blubba

368 TABLE CAPTIONS

369 TABLES

R Pack- ages	Version	Description
ggplot2	https://cran.r-project.org/web/packages/ggplot2/index.html - ggplot2_3.3.0 or later	ggplot2 is a system for declaratively creating graphics, based on The Grammar of Graphics.
NMF	https://cran.r-project.org/web/packages/NMF/index.html - NMF_0.22.0 or later	Provides a framework to perform Non-negative Matrix Factorization (NMF).

Table 2: **R Package dependencies** blubba