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]

[1]Klaus Tschira Institute for Integrative Computational Cardiology, University Heidelberg, 69120 Heidelberg, Germany
[2]Department of Internal Medicine III (Cardiology, Angiology, and Pneumology), University Hospital Heidelberg, 69120 Heidelberg, Germany
[3]German 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

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

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

MATERIALS

ONT direct RNA sequencing

1. 500 ng polyA$^+$ RNA isolated from total RNA e.g. with Oligotex mRNA kit (Qiagen) or Dynabeads oligo dT_{25} beads (Thermo Fisher Scientific) or *in vitro* transcriptome sample. Store RNA at -80 °C and the mRNA purification kit as recommended by the manufacturer.

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

Preparation of an *in vitro* transcriptome sample

1. 100 ng polyA$^+$ RNA isolated from total RNA e.g. with Oligotex mRNA kit (Qiagen) or Dynabeads oligo dT$_{25}$ beads (Thermo Fisher Scientific). Store RNA at -80 °C and the mRNA purification kit as recommended by the manufacturer
2. 10 μM oligo(dT)-VN RT primer. TTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN. Store at -20 °C.
3. 20 μM template switching oligo (TSO). ACTCTAATACGACTCACTATAGGGAGAGGGCrGrG+G. Store at -20 °C.

4. 10 μM T7 extension primer. GCTCTAATACGACTCACTATAGG. Store at -20 °C.

5. Nuclease-free water. Store at room temperature.

6. dNTP Mix (10 mM each). Store at -20 °C.

7. Template Switching RT Enzyme Mix (New England Biolabs). Store at -20 °C.

8. Q5 Hot Start High-Fidelity 2X Master Mix (New England Biolabs). Store at -20 °C.

9. RNase H (5,000 U/ml) (New England Biolabs). Store at -20 °C.

10. NucleoSpin Gel and PCR Clean-up, Mini kit for gel extraction and PCR clean up (Macherey-Nagel) or equivalent. Store at room temperature.

11. MEGAscript T7 transcription kit (Thermo Fisher Scientific). Store at -20 °C.

12. RNA Clean & Concentrator-25 kit (Zymo Research). Store at room temperature.

13. Thermocycler.

14. Table top centrifuge for 1.5 ml tubes.

15. Nanodrop spectrophotometer or equivalent.

16. 0.2 ml PCR tubes, 1.5 ml DNA LoBind tubes (Eppendorf).

Hardware requirements

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

Software dependencies and installation

Our analysis workflow has few requirements, which are detailed in Table 2. Specifically, to execute our workflow, the following prerequisites are necessary: a BASH shell, a JAVA runtime environment, a working PERL and R installation. Additional i.e. non-standard software to process and map Nanopore reads (bedtools, samtools and Minimap2) are obligatory, while the installation of a Nanopore read simulator (NanoSim) is optional and depends on your use case. Table ?? lists some additional R packages, which are required to run the R code. Detailed instructions on how to setup are found under `https://github.com/dieterich-lab/MiMB_JACUSA2_chapter`

METHODS

Overview Figure 1

Nanopore direct RNA sequencing

1. Adjust 500 ng polyA$^+$ RNA to a total volume of 9 μl with nuclease-free water. Complete RT adapter ligation reaction (in 0.2 ml PCR tube) with 3 μl NEBNext Quick Ligation Reaction Buffer, 0.5 μl RNA CS (RCS, from SQK-RNA002), 1 μl RT-Adapter (RTA, from SQK-RNA002) and 1.5 μl T4 DNA Ligase. Incubate 10 min at room temperature.

2. Prepare reverse transcription master mix on ice during ligation: 9 μl nuclease-free water, 2 μl 10 mM dNTPs, 8 μl 5x SuperScript IV first strand buffer, 4 μl 0.1 mM DTT.

3. Add the reverse transcription master mix to the ligation reaction and mix by pipetting. Add 2 μl SuperScript IV reverse transcriptase and mix by pipetting. Incubate in a thermocycler with the following protocol: 50 min at 50 °C, 10 min at 70 °C, cool down to 4 °C.

4. Let the Agencourt RNAClean XP beads come to room temperature during reverse transcription. Carefully resuspend beads before use. Transfer reaction to a 1.5 ml DNA LoBind tube and mix with 72 μl Agencourt RNAClean XP beads. Incubate 5 min at room temperature on a gentle rotator mixer.

5. Collect beads on a magnetic stand and remove supernatant. Wash pelleted beads two times (30 sec) with 200 μl freshly prepared 70 % ethanol. Remove supernatant. Spin sample down and place on magnet again. Remove any residual ethanol.

6. Resuspend beads in 20 μl nuclease-free water by gentle flicking and incubate 5 min at room temperature on a gentle rotator mixer. Collect beads on a magnetic stand and transfer 20 μl eluate in a fresh 1.5 ml DNA LoBind tube.

7. For ligation of the RMX adapter, add the following to 20 μl eluate: 8 μl NEBNext Quick Ligation Reaction Buffer, 6 μl RMX (from SQK-RNA002), 3 μl nuclease-free water, 3 μl T4 DNA Ligase. Mix by pipetting and incubate 10 min at room temperature.

8. Add 40 μl carefully resuspended Agencourt RNAClean XP beads to the reaction and mix by pipetting. Incubate 5 min at room temperature on a gentle rotator mixer.

9. Collect beads on a magnetic stand and remove supernatant. Wash pelleted beads two times with 150 μl wash buffer (WSB, from SQK-RNA002). Resuspend beads by flicking, spin down and return to magnetic stand. Remove supernatant from pelleted beads.

10. Resuspend beads in 21 μl elution buffer (EB, from SQK-RNA002) by gentle flicking and incubate 5 min at room temperature on a gentle rotator mixer. Pellet beads on a magnetic stand and transfer 21 μl eluate in a fresh 1.5 ml DNA LoBind tube.

11. Quantify 1 μl of the library on a Qubit fluorometer with the Qubit dsDNA HS kit according to the manufacturerers protocol. Concentration should be usually in the range of 5 - 10 ng/μl.

12. Insert MinION R9.4.1 Flow cell in the MinION or GridION sequencing device and perform Flow cell check in the MinKNOW software. For successful sequencing of mammalian polyA$^+$ RNA at least 1,000 available pores are recommended.

13. Prepare Priming Mix by adding 30 μl flush tether (FLT, from EXP-FLP002) to a vial of flush buffer (FB, from EXP-FLP002) and mix by pipetting. Open priming port. Remove air bubble from priming port by inserting the tip of a P1000 pipette into the priming port and slowly dialing up, until a small volume of storage buffer enters the pipette tip. Load 800 μl Priming Mix via the priming port and carefully avoid introduction of air bubbles. Close the priming port and wait for 5 min.

14. Mix 20 μl library with 17.5 μl nuclease-free water and 37.5 μl RNA running buffer (RRB, from SQK-RNA002) and mix by pipetting. Open the priming port and the sample port. Load 200 μl Priming Mix via the priming port. Mix library by pipetting just before loading and load dropwise via the sample port. Carefully avoid introduction of air bubbles. Close the sample port and the priming port.

15. Start sequencing for 48 to 72 h in the MinKNOW software. Choose direct RNA-sequencing kit and high-accuracy basecalling as parameters. We recommend to adjust the output filter to a minimum Q score of 7 (instead of 9).

Preparation of an *in vitro* transcriptome sample

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

1. Adjust 100 ng polyA^{+} RNA to a total volume of 6 μl with nuclease-free water. Add 1 μl each of 10 μM oligo(dT)-VN RT primer and 10 mM dNTPs. Mix by pipetting and incubate in a thermocycler: 5 min at 75 °C, 2 min at 42 °C, cool to 4 °C.

2. Assemble 2.5 μl 4x template switching RT buffer, 0.5 μl 20 μM TSO, 1 μl 10x template switching RT enzyme mix and mix by pipetting. Combine with 6 μl RNA and incubate in a thermocycler: 90 min at 42 °C, 10 min at 68 °C, cool to 4 °C.

3. For Second strand synthesis add to First strand synthesis reaction: 50 μl Q5 Hot Start High-Fidelity 2X Master Mix, 5 μl RNase H, 2 μl 10 μM T7 extension primer, 33 μl nuclease-free water. Mix by pipetting and incubate in a thermocycler: 15 min at 37 °C, 1 min at 95 °C, 10 min at 65 °C, cool to 4 °C.

4. Purify double stranded cDNA with NucleoSpin Gel and PCR Clean-up kit according to the manufacturerers protocol and elute in 20 μl elution buffer. Determine concentration on a Nanodrop spectrophotometer. cDNA may be stored at -20 °C.

5. Combine 8 μl cDNA for *in vitro* transcription with 2 μl each of ATP, GTP, CTP, UTP, 10x reaction buffer and enzyme mix from the MEGAscript T7 transcription kit. Incubate 3 h at 37 °C.

6. Digest template DNA by addition of 1 μl Turbo DNase. Mix by pipetting and incubate 15 min at 37 °C.

7. Adjust reaction volume to 100 μl with nuclease-free water and clean up with RNA Clean & Concentrator-25 kit according to the manufacturers protocol, using two volumes of adjusted RNA binding buffer (1:1 RNA binding buffer : ethanol). Elute RNA in 25 μl nuclease-free water. Determine RNA concentration on a Nanodrop spectrophotometer. Store at -80 °C.

Nanopore read processing

1. first steps

2. JACUSA2 requires sorted and indexed BAM files. To sorte and create a BAM file index use the following SAMtools commands.

```
$ samtools sorte mapping.bam mapping.sorted.bam
$ samtools index mapping.sorted.bam
```

3. Add the MD tag field to the BAM files. This requires the reference sequence 'reference.fasta' used for the mapping. The MD tag field stores information on mismatched and deleted reference bases.

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

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

The benchmark is composed of two samples from two conditions: wild-type (positive condition) and knock-out(control condition) for Hek293 cell line with two replicates. Given the preprocessed mapped reads as input (BAM files) 'HEK293T-WT-rep2.bam, HEK293T-WT-rep3.bam representing the wildtype replicates and HEK293T-KO-rep2.bam and HEK293T-KO-rep3.bam as the control replicates,

1. Identify read error profile: run JACUSA2 with paired samples option (call-2). One should specify the corresponding library information and may filter reads according to many parameters. Here, we consider all positions with read coverage > 4. The output consists of read error profile where the format is a combination of BED6 with JACUSA2 methods specific columns and common info columns: "info", "filter", and "ref". Check JACUSA2 manual for more details on JACUSA2 filter and output options.

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

2. Preprocess JACUSA2 output: given the JACUSA2 output, select non-overlapping sites of homo-polymer regions (JACUSA filter: Y) and within a 5mer of a central A nucleotide flanked by 2 adjacent random nucleotides (NNANN). Each site is represented by the insertion, deletion and mismatch scores and the position number within the 5mer specific context. The 'README_processing.sh' bash script performs the preprocessing and produces a text file 'call2_SitesExt2_indel_slim2.txt' containing the features for the selected sites and a separate file representing the 5mer bases 'checkMotif_reformat.txt'. One may precise the path to outputs within the command.

```
$ bash README_processing.sh WT_vs_KO_RC22_call2_result.out hg38.genome GRCh38_9
```

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

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

4. Extract m6A modification patterns: now that one got the matrix of Sites X Features, the next step is to extract patterns allowing to predict m6A modified positions. To this end, the non-negative matrix factorization (NMF) analysis is suggested. The R script 'HEK293_data_prep_step2.R' allows generating patterns from a subset of the data associated to previously reported m6A sites. Here, the unsupervised pattern learning is based on 2401 m6A sites (). Based on the silhouette and cophenetic correlation indices, we could identify an optimal factorization rank of 7 (fig. 4). reference

```
$ Rscript HEK293_data_prep_step2.R path_to_output miCLIP_union_flat_exclude_Y_c
```

the 'miCLIP_union_flat_exclude_Y_chromosome.bed' file contains all m6A sites reported in (). reference

5. Predict m6A modifications: the empirical Cumulative Distribution Function (eCDF) of the detected patterns scores can be used to predict novel m6A sites. We examine the ability of prediction on a subset of ... reported m6A sites and we plot the eCDF of pattern 3 scores by category (fig. 5). The R script 'HEK293_data_prep_step3.R' allows generating the eCDF probabilities of modification as fellow.

```
$ Rscript HEK293_data_prep_step3.R path_to_output miCLIP_union_flat_exclude_Y_c
```

Use Case 2: Comparison of wildtype and IVT samples

Christoph

Use Case 3: Comparison of wildtype to simulated IVT sample

In case the control condition is not available, NanoSim tool can generate *in silico* synthetic sample from the reference genome as a control condition. Then, use the simulated read for JACUSA2 paired conditions analysis.

1. Generate *in silico* synthetic sample: the first step to generate *in silico* reads is read characterization, which produces a set of read profiles serving as the input to the next step, the simulation stage. For more details check NanoSim manual on. We generate reads in genome mode, which takes a reference genome and a training read set in FASTA or FASTQ format as input:

```
$ read_analysis.py genome -i data.fasta -ga G_ALNM -o HEK293_
```

Then, the simulation stage (in a genome mode) takes reference genome and read profiles as input and outputs simulated reads in FASTA format.

```
$ simulator.py genome -rg REF_G  -c HEK293_char -o HEK293_sim
```

after getting the *in silico* synthetic sample, the next steps are similar to the first cases.

2. Identify read error profile: given the *in silico* synthetic sequence, one can run JACUSA2 on paired conditions mode with the same parameters as the previously described cases.

```
$ JACUSA2 2.0.0-RC22 call-2 -m 1 -q 1 -c 4 -p 10 -D -I -a D,Y -P1 FR-SECONDSTRA
-P2 FR-SECONDSTRAND -r WT_vs_KO_2samp_RC22_call2_result.out HEK293T-WT-rep2.bam
HEK293T-IVT-rep1.bam,HEK293T-IVT-rep2.bam
```

3. Preprocess JACUSA2 output: select the 5mer specific sites (NNANN) considering the Y filter as follows.

```
$ bash README_processing.sh WT_vs_IVT_RC22_call2_result.out hg38.genome GRCh38_
```

4. Extract 5mer features using the following command:

```
$ Rscript Code/HEK293_data_prep.R path_to_output WT_vs_IVT_RC22_call2_result.ou
```

5. Extract m6A modification patterns based on 2401 m6A sites (). From the silhouette and cophenetic correlation indices, we could identify an optimal factorization rank of 7 (fig. 6). reference

```
$ Rscript HEK293_data_prep_step2.R path_to_output miCLIP_union_flat_exclude_Y_c
```

6. Predict m6A modifications: we plot the eCDF of pattern 7 scores by category (fig. 7)

```
$ Rscript HEK293_data_prep_step3.R path_to_output miCLIP_union_flat_exclude_Y_c
```

NOTES

Tips and Tricks

ACKNOWLEDGMENTS

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

REFERENCES

Samir Adhikari, Wen Xiao, Yong-Liang Zhao, and Yun-Gui Yang. m(6)a: Signaling for mrna splicing. *RNA biology*, 13:756–759, September 2016. ISSN 1555-8584. doi: 10.1080/15476286.2016.1201628.

Ina Anreiter, Quoseena Mir, Jared T. Simpson, Sarath C. Janga, and Matthias Soller. New twists in detecting mrna modification dynamics. *Trends in biotechnology*, 39:72–89, January 2021. ISSN 1879-3096. doi: 10.1016/j.tibtech.2020.06.002.

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

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

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

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

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

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

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

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

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

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

Guanqun Zheng, John Arne Dahl, Yamei Niu, Peter Fedorcsak, Chun-Min Huang, Charles J. Li, Cathrine B. Vågbø, Yue Shi, Wen-Ling Wang, Shu-Hui Song, Zhike Lu, Ralph P. G. Bosmans, Qing Dai, Ya-Juan Hao, Xin Yang, Wen-Ming Zhao, Wei-Min Tong, Xiu-Jie Wang, Florian Bogdan, Kari Furu, Ye Fu, Guifang Jia, Xu Zhao, Jun Liu, Hans E. Krokan, Arne Klungland, Yun-Gui Yang, and Chuan He. Alkbh5 is a mammalian rna demethylase that impacts rna metabolism and mouse fertility. *Molecular cell*, 49:18–29, January 2013. ISSN 1097-4164. doi: 10.1016/j.molcel.2012.10.015.

FIGURE CAPTIONS

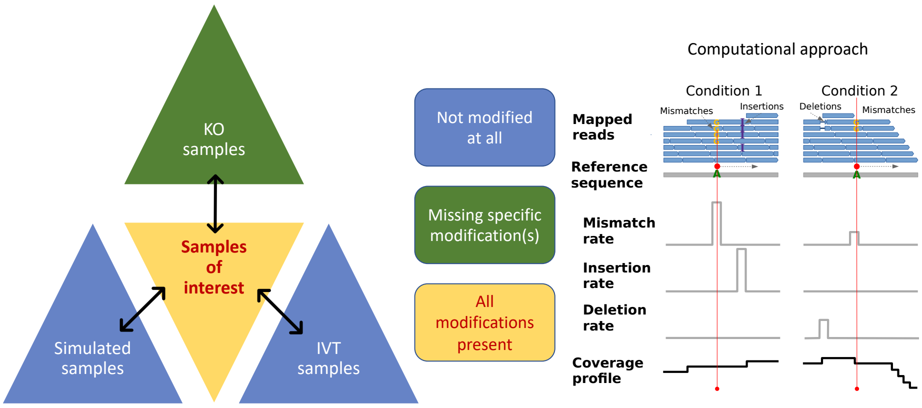

Figure 1: **General outline of RNA modification detection by JACUSA2.** A key feature of our approach is that multiple replicates can be compared as shown on the left. Samples of interests where all modifications are present could be compared with either KO samples where the modification of interest 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.

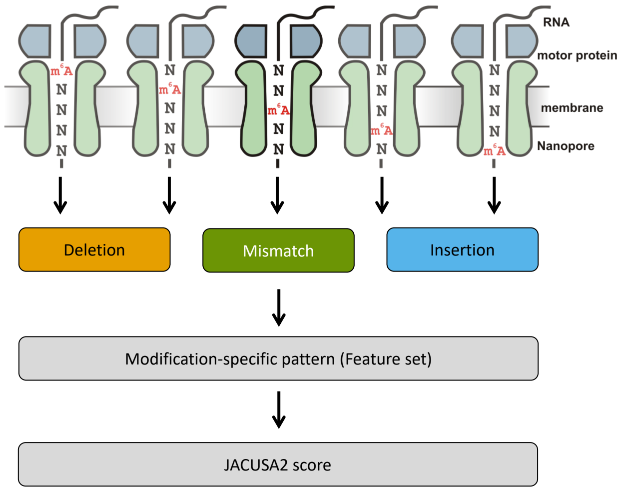

Figure 2: **Motivation of 5mer context for RNA modification mapping**. The nanopore covers 5 consecutive RNA residues. That is why we consider a 5mer context and derive 3 principal features for every position within a given 5 mer (15 features in total, with a central A residue in this example). We evaluate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

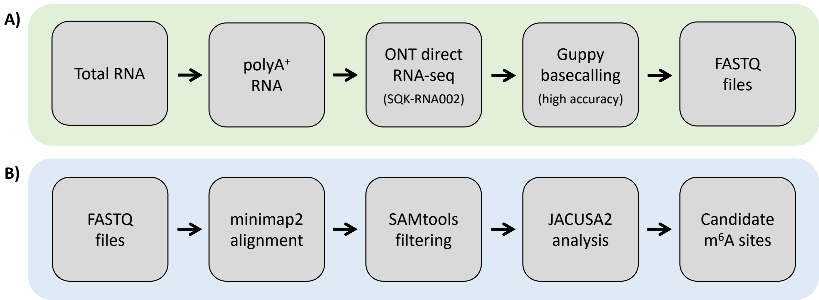

Figure 3: **Experimental and computational workflow**. tbd

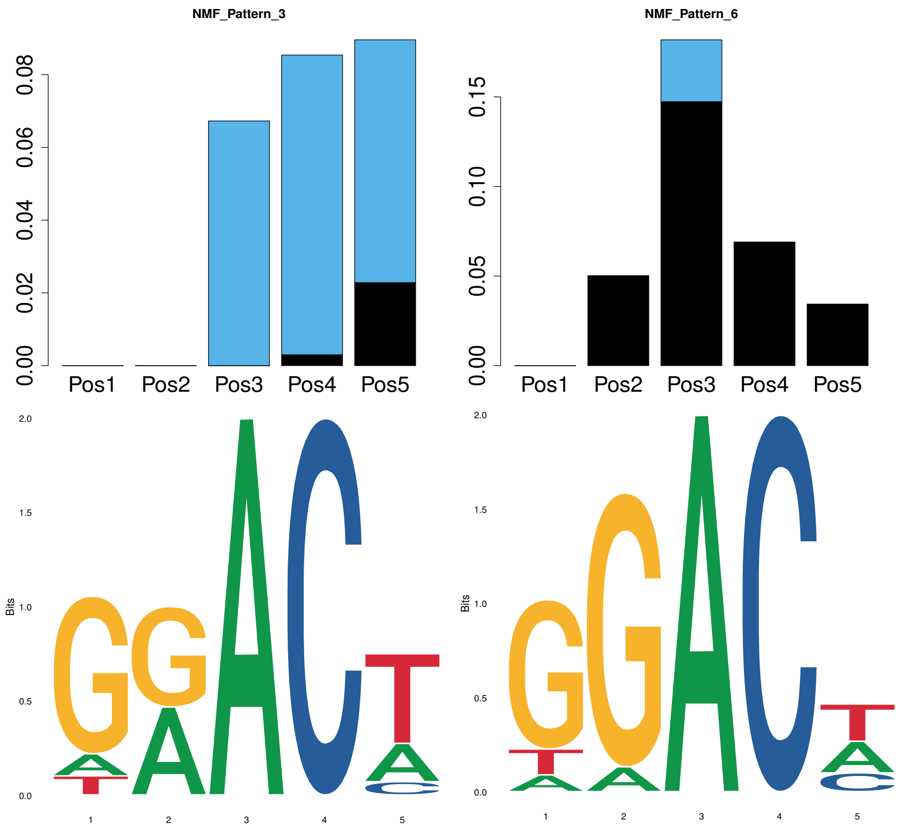

Figure 4: **Main patterns for WT vs KO case.**

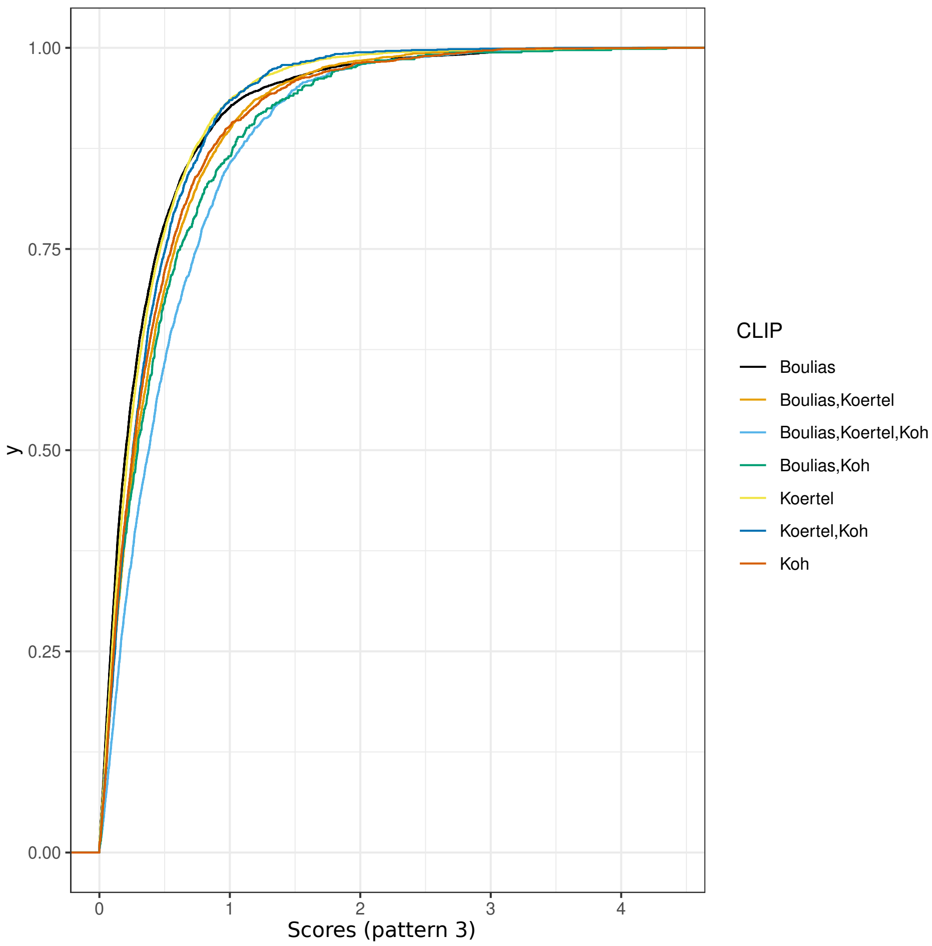

Figure 5: **ecdf for WT vs KO case.**

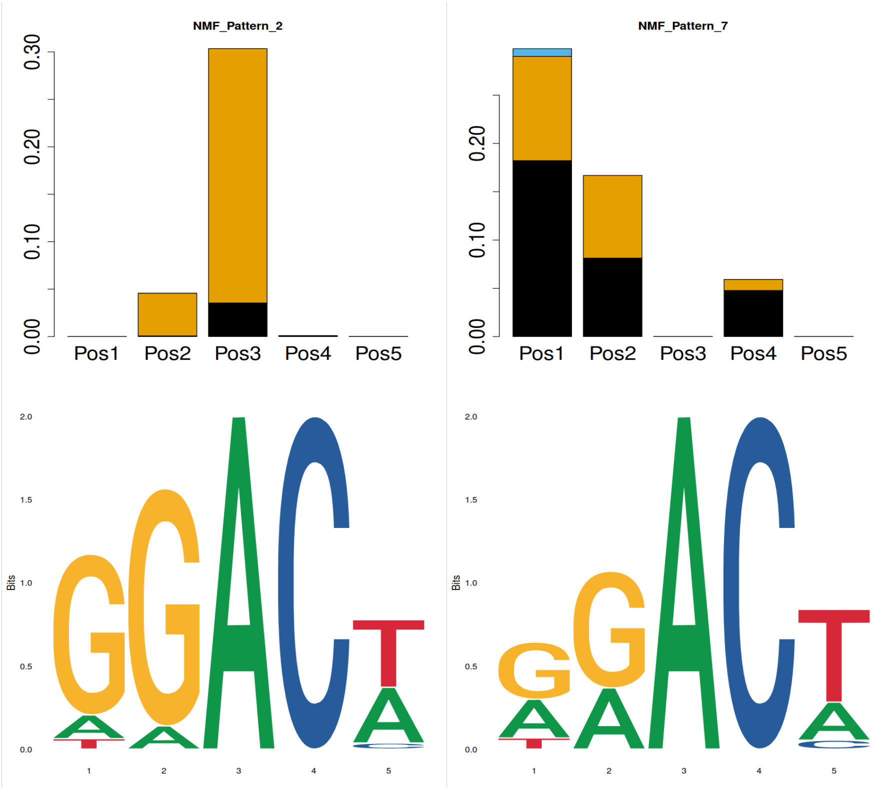

Figure 6: **Main patterns for WT vs IVT case.**

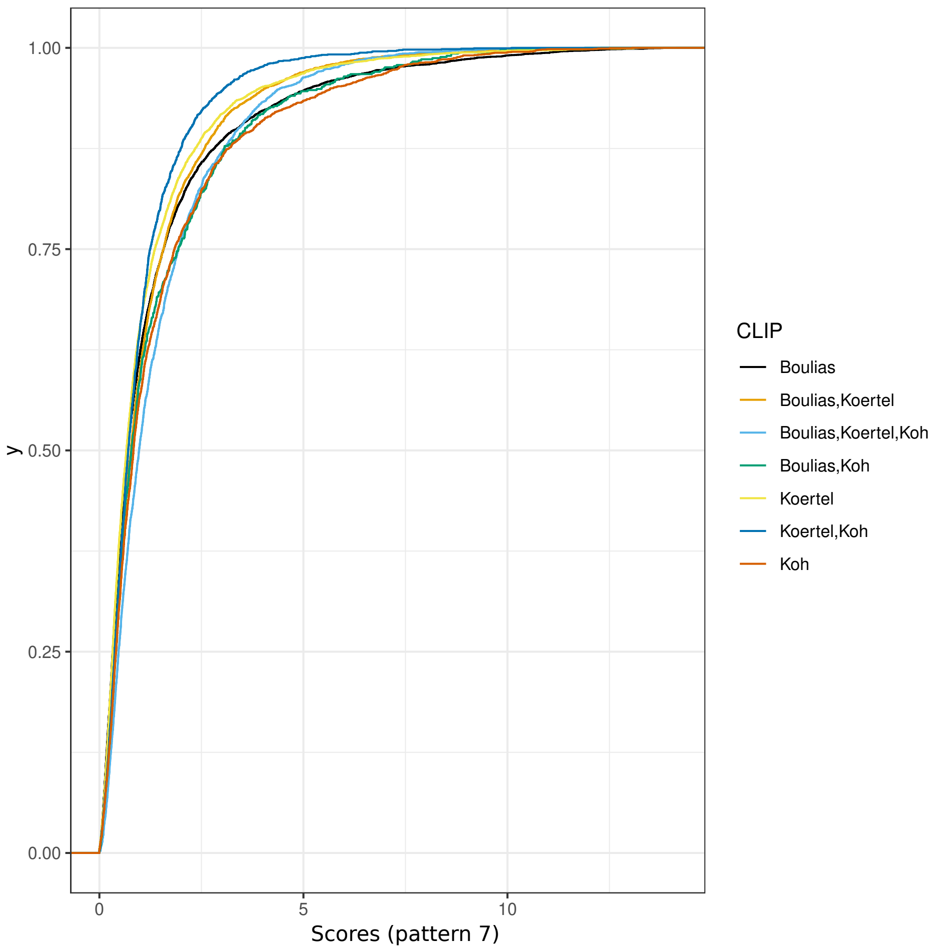

Figure 7: **ecdf for WT vs IVT case.**

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

TABLE CAPTIONS

TABLES

R Packages	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