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 [**?**]. 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 **???**. The modification is catalyzed co-transcriptionally by a Mega-Dalton methyltransferase complex, which includes the heterodimer METTL3-METTL14 and other associated subunits **?**. This modification is reversible since two proteins of the AlkB-family demethylases can remove m6A from mRNA transcripts [**??**]. 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 [**???**]. 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 [**??**]. 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. TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN. Store at -20 °C.
3. 20 μM template switching oligo (TSO). ACTCTAATACGACTCAC-TATAGGGAGAGGGCrGrG+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.* **?** 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

Minimap2 and samtools

Christoph

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

Xpore data

Christoph

Use Case 2: Comparison of wildtype and IVT samples

Christoph

Use Case 3: Comparison of wildtype to simulated IVT sample

Christoph

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.

FIGURE CAPTIONS

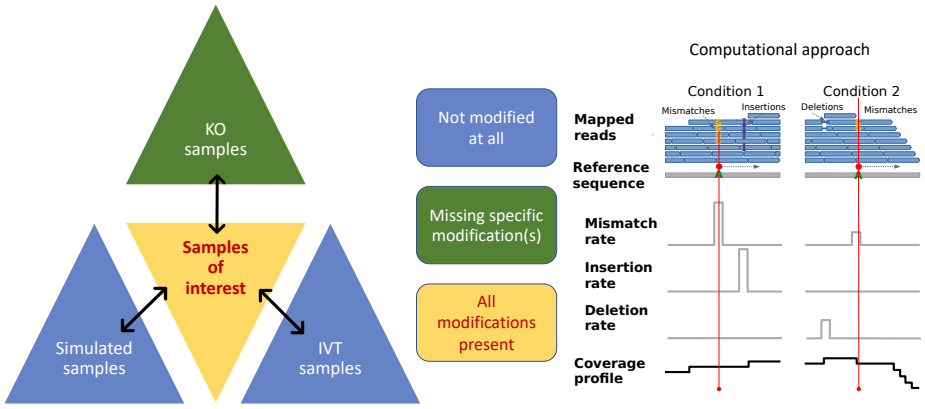

Figure 1: **General outline of RNA modification detection by JACUSA2.** A key feature of our approach is that multiple replicates can be compared as shown on the left. Samples of interests where all modifications are present could be compared with either KO samples where the modification of interest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

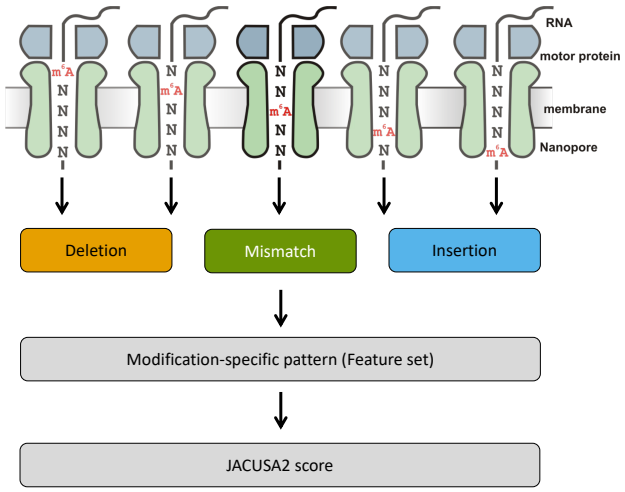

Figure 2: **Motivation of 5mer context for RNA modification mapping**. The nanopore covers 5 consecutive RNA residues. That is why we consider a 5mer context and derive 3 principal features for every position within a given 5 mer (15 features in total, with a central A residue in this example). We evaluate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

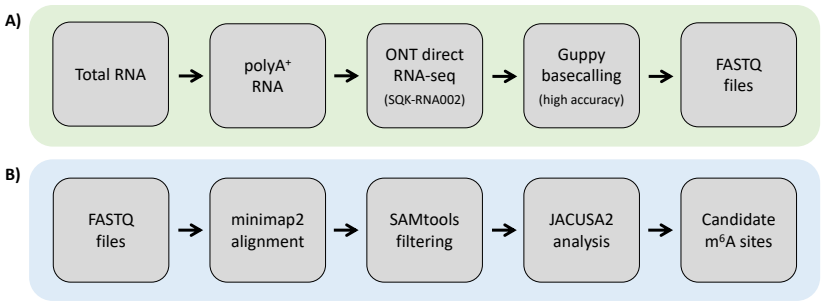

Figure 3: **Experimental and computational workflow**. tbd

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