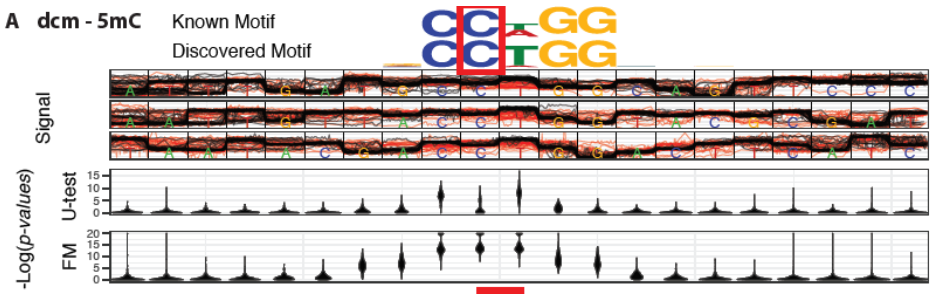
RNA modification identification using nanopore direct RNA sequencing

***biorxiv*, April 2017**

De novo Identification of DNA Modifications Enabled by Genome-Guided Nanopore Signal Processing

Authors: Marcus Stoiber1, Joshua Quick2, Rob Egan3, Ji Eun Lee3, Susan Celniker1, Robert K. Neely4, Nicholas Loman2, Len A Pennacchio1,3, James Brown1,5,6,7

- **Tombo/nanoraw: https://nanoporetech.github.io/tombo/**
- resquiggle the genome to associate raw nanopore signal with genomic positions and unsupervised statistical testing
- nomalized raw signal
- ✓ no need for prior external or additional training data
- ✓ effective visualization of raw nanopore signal in genomic contexts
- X modification type depends on the non-modified control data
- X aligner not splice-aware

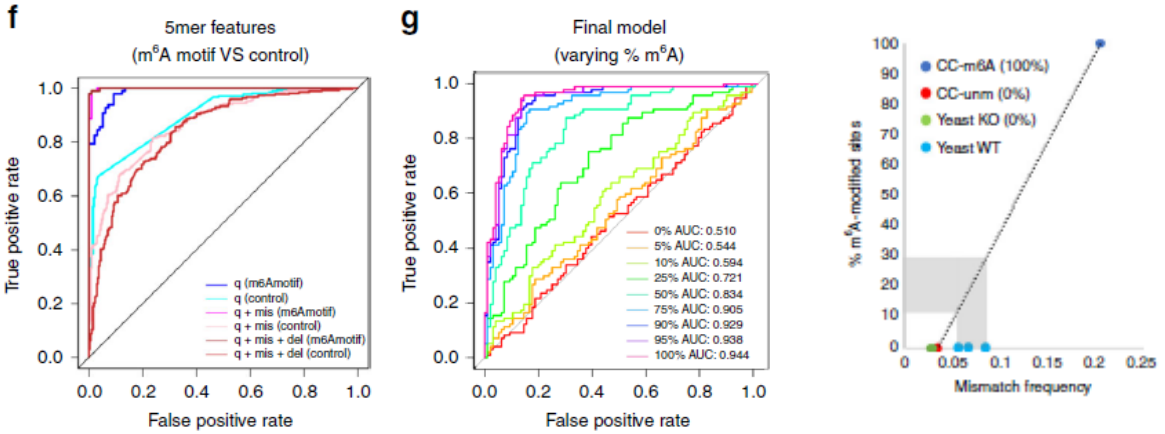


***Nature Communication*, 2019 Sep**

Accurate detection of m^{6}A RNA modifications in native RNA sequences

Huanle Liu[1,2,11], Oguzhan Begik[1,2,3,11], Morghan C. Lucas[1,4], Jose Miguel Ramirez[1], Christopher E. Mason[5,6,7], David Wiener[8], Schraga Schwartz[8], John S. Mattick[2,3,10], Martin A. Smith[3,9] & Eva Maria Novoa[1,2,3,4]

- **EpiNano: https://github.com/enovoa/EpiNano**
- SVM model trained on IVT comprised all possible 5-mers using canonical A or m6A
- base-called “errors” information: base quality, deletion frequency and mismatch frequency
- ✓ reasonable accuracy at more than 25% of methylation ratio
- ✓ linear regression to estimate the stoichiometry
- ✓ DRS of polyA(+) from *S. cerevisiae wt* and *ime4*Δ
- X limited to m6A modification
- X not used to DRACH k-mer with more than one A

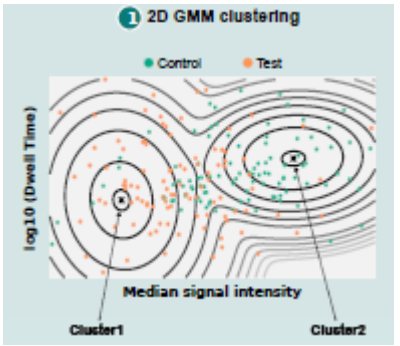


***biorxiv*, 2019 Nov**

RNA modifications detection by comparative Nanopore direct RNA sequencing

Adrien Leger[1*], Paulo P. Amaral[2,3*], Luca Pandolfini[2], Charlotte Capitanchik[4], Federica Capraro[4,5], Isaia Barbieri[2,8], Valentina Migliori[2], Nicholas M. Luscombe[4,6,7], Anton J Enright[9], Konstantinos Tzelepis[2], Jernej Ule[4,5], Tomas Fitzgerald[1], Ewan Birney[1**], Tommaso Leonardi[2,10**] and Tony Kouzarides[2**]

- **Nanocompore: https://github.com/tleonardi/nanocompore**
- various statistical test: Kolmogorov-Smirnov (KS),Mann-Whitney (MW) and Welch's t-test.
- median signal intensity and the log10(dwell time)
- single-molecule modification probabilities from GMM clustering
- ✓ does not require a training set
- ✓ allows replicates to model biological variability
- ✓ single molecule resolution, modification stoichiometry and combinatorics.
- ✓ applied as-is to any RNA modification
- ✓ DRS and miCLIP of polyA(+) from *METTL3* KD and *WT* MOLM13 cells
- X non-modified control data required: knock-downs, knock-outs, cDNA or IVT samples
- X modification type depends on the control data
- X unsuitable for the identification of very low frequency modifications

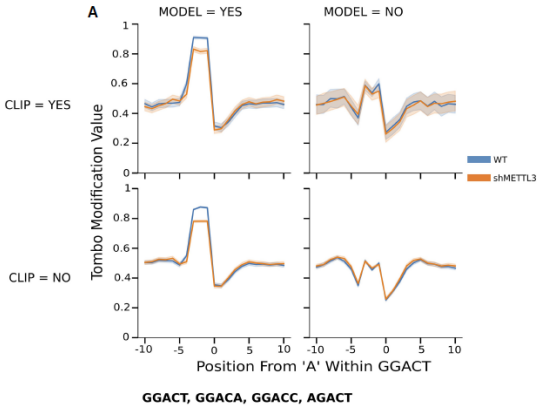


***RNA*, 2020 Jan**

Direct RNA sequencing enables m^{6}A detection in endogenous transcript isoforms at base-specific resolution

DANIEL A. LORENZ,[1,2,3,4] SHASHANK SATHE,[1,2,3,4] JACLYN M. EINSTEIN,[1,2] and GENE W. YEO[1,2,3]

- **MINES: https://github.com/YeoLab/MINES**
- random forest classifier trained on known miCLIP sites
- fraction modification values calculated by tombo
- a separate model for each 5mer within the DRACH motif
- ✓ single site, isoform-level resolution
- ✓ DRS of polyA(+) from *HEK-WT, HEK-shMETTL3, HMEC-WT, HMEC-ALKBH5*
- X limited to m6A sites within 4 specific DRACH sequences with AUC values >0.67
- X affected by the same biases and/or limitations as miCLIP

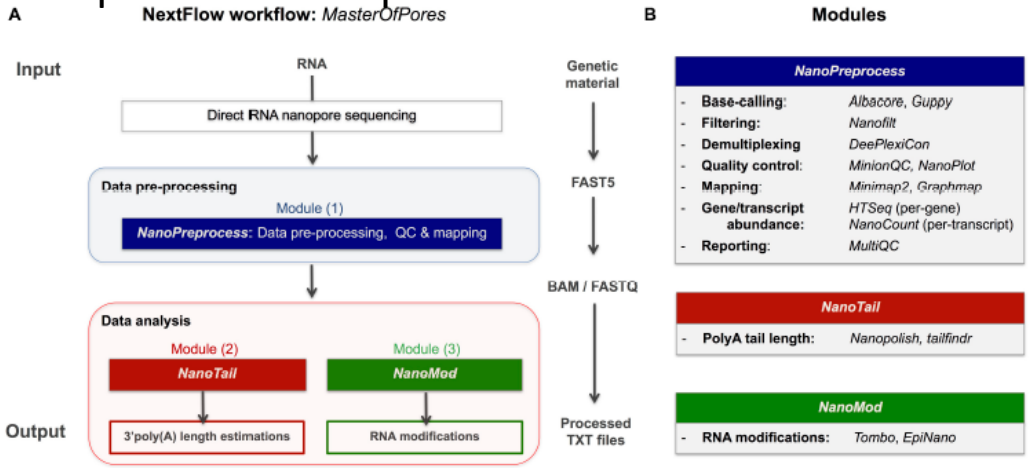


frontiers in Genetics, 2020 Mar

MasterOfPores: A Workflow for the Analysis of Oxford Nanopore Direct RNA Sequencing Datasets

Luca Cozzuto[1], Huanle Liu[1], Leszek P. Pryszcz[1,2], Toni Hermoso Pulido[1], Anna Delgado-Tejedor[1,3], Julia Ponomarenko[1,3] and Eva Maria Novoa[1,3,4,5]**

- **MasterOfPores: https://github.com/biocorecrg/master_of_pores**
- NextFlow framework to construct a standadized workflow to analyze DRS data
- pre-processing and data analysis modules
- one MinION run takes ~2h on a CPU cluster using 100 nodes and ~1h on a single GPU
- ✓ workflow management systems together with Linux containers
- ✓ highly reproducible, scalable and parallelizable
- ✓ support of differen batch schedulers, cloud platforms, GPU computing
- X requires two samples to detect RNA modifications

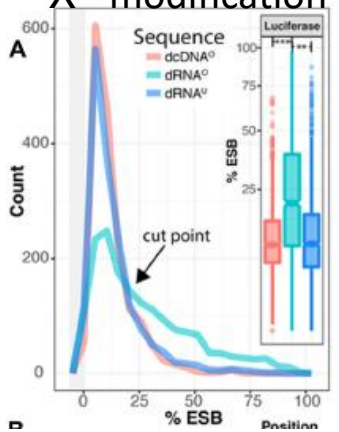


NAR, 2020 July

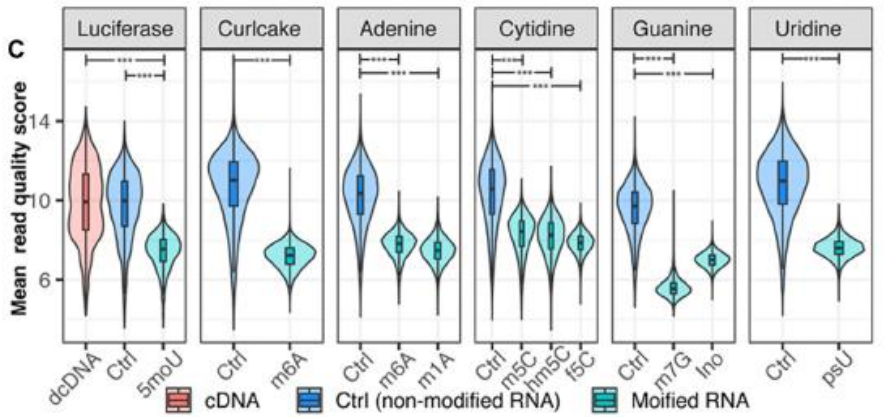
Decoding the epitranscriptional landscape from native RNA sequences

Piroon Jenjaroenpun[1], Thidathip Wongsurawat[1], Taylor D. Wadley[1], Trudy M. Wassenaar[2], Jun Liu[3], Qing Dai[3], Visanu Wanchai[1], Nisreen S. Akel[4], Azemat Jamshidi-Parsian[5], Aime T. Franco[4], Gunnar Boysen[6], Michael L. Jennings[4], David W. Ussery[1], Chuan He[3] and Intawat Nookaew[1,4,*]

- **ELIGOS: https://gitlab.com/piroonj/eligos2**
- Fisher's exact test between the native RNA and a reference (IVT RNA, unmodified in vitro transcription RNA or a background error model (rBEM))
- percent Error of Specific Bases (%ESB) differences
- IVT of human mRNA to construct rBEMs of all pentamers
- ✓ sequence context, background errors, homopolymeric sequence
- ✓ m6A, m1A, m5C, hm5C, f5C, psU, m7G, Ino
- ✓ DRS and cDNA of IVT comprised all possible 5-mers using modified base or non-modified base such as B5(NNNNN containing at least one A)B5
- ✓ DRS of *Mettl3* KO, *Mettl14* KO and control mESCs, various rRNA
- X modification type not specified, not used to m5C



High %ESB values more frequently obtained with the native RNA



***biorxiv*, 2021 Jan**

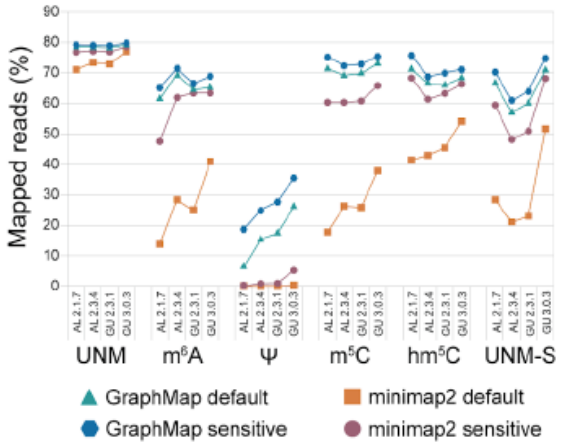
Quantitative profiling of native RNA modifications and their dynamics using nanopore sequencing

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

- **nanoRMS: https://github.com/novoalab/nanoRMS**
- base-calling error features to identify the modified sites, KNN and k-means to bin read into two clusters accroding to per-read current intensity/trace features

- ✓ pseudouridine and 2'-O-methylation
- ✓ quantitative identification, even at very low modification stoichiometriess
- ✓ support RNA modification stoichiometry estimation using both nanopolish and tombo resquiggling
- ✓ systematically analyze the features for different modifications and mapping tools

- X base-calling errors features are not suitable for all modifications and sequence context
- X detection low stoichiometry was only possible when using comparison of pairwise conditions.

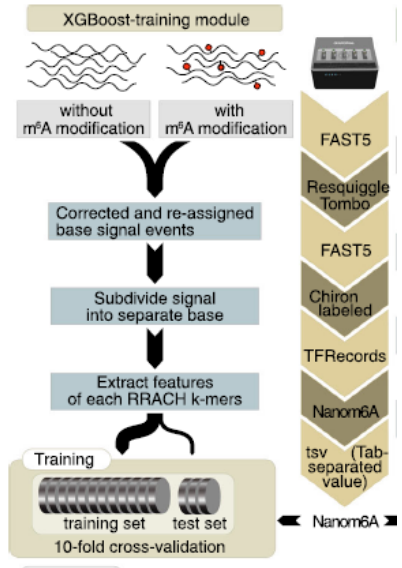


***Genome Biology*, 2021 Jan**

Quantitative profiling of N^6-methyladenosine at single-base resolution in stem-differentiating xylem of *Populus trichocarpa* using Nanopore direct RNA sequencing

Yubang Gao[1†], Xuqing Liu[2†], Bizhi Wu[2], Huihui Wang[2], Feihu Xi[1], Markus V. Kohnen[1], Anireddy S. N. Reddy[3] and Lianfeng Gu[2*]

- **Nanom6A: https://github.com/gaoyubang/nanom6A**
- XGBoost model trained on Epinano IVT dataset
- mean, standard deviation, median, and dwell time of each read
- ✓ Compare with published tools on several datasets
- ✓ quantitative profiling of m6A in stem-differentiating xylem of Populus trichocarpa, 80% m6A sites validated by MeRIP-Seq and m6A-REF-seq
- X limited to m6A RRACH k-mers



***Nature Methods*, 2019 Dec**

Nanopore native RNA sequencing of a human poly(A) transcriptome

Rachael E. Workman[1,9], Alison D. Tang[2,3,9], Paul S. Tang[4,9], Miten Jain[2,3,9], John R. Tyson[5,9], Roham Razaghi[1,9], Philip C. Zuzarte[4], Timothy Gilpatrick[1], Alexander Payne[6], Joshua Quick[7], Norah Sadowski[1], Nadine Holmes[6], Jaqueline Goes de Jesus[7], Karen L. Jones[5], Cameron M. Soulette[2,3], Terrance P. Snutch[5], Nicholas Loman[7], Benedict Paten[2,3], Matthew Loose[6], Jared T. Simpson[4,8], Hugh E. Olsen[2,3,10], Angela N. Brooks[2,3,10], Mark Akeson[2,3,10]* and Winston Timp[1,10]*

- DRS and cDNA seq of polyA(+) from GM12878, 30 flow cells in 6 laboratories, 9.9 million aligned reads
- Isoform detection and analysis, allele-specific expression, poly(A) tail length, m6A and RNA editing

***eLife*, 2020 Jan**

Nanopore direct RNA sequencing maps the complexity of Arabidopsis mRNA processing and m^{6}A modification

Matthew T Parker[1†], Katarzyna Knop[1†], Anna V Sherwood[1†‡], Nicholas J Schurch[1†§], Katarzyna Mackinnon[1], Peter D Gould[2], Anthony JW Hall[3], Geoffrey J Barton[1]*, Gordon G Simpson[1,4]*

[1]School of Life Sciences, University of Dundee, Dundee, United Kingdom; [2]Institute of Integrative Biology, University of Liverpool, Liverpool, United Kingdom; [3]Earlham Institute, Norwich Research Park, Norwich, United Kingdom; [4]James Hutton Institute, Invergowrie, United Kingdom

- DRS and Illumina seq of polyA(+) and 5'-capped mRNAs from Arabidopsis wt, *vir-1* mutant and VIR-GFP complemented seedlings, ERCC RNA spike-in
- transcriptoin start site, novel splice sites and isoforms, m6A modifications, alternative polyadenylation site, poly(A) tail length

***Genome Research*, 2020 Feb**

Direct full-length RNA sequencing reveals unexpected transcriptome complexity during *Caenorhabditis elegans* development

Runsheng Li,[1,3] Xiaoliang Ren,[1,3] Qiutao Ding,[1] Yu Bi,[1] Dongying Xie,[1] and Zhongying Zhao[1,2]

- DRS of polyA(+) of *C. elegans* embryo (EMB), L1 larva (L1), and young adult (YA)
- full-length reads
- TrackCluster: novel isoform identification and quantification
  - ✓ alternative use of promoter or polyadenylation sites
  - ✓ UTR extensions or truncations at the 5′ or 3′ end
  - ✓ new combinations of exons within the gene body
  - ✓ intron retention
  - ✓ isoform fusion
- RNA modifications