Identifying NSUN2- and NSUN6- mediated mRNA m^{5}C modification from the noise

Jianheng Liu[1], Tao Huang[1], Wanying Chen[1], Yusen Zhang[1], Tianxuan Zhao[1], Rui Zhang[1,2,3]

1. MOE Key Laboratory of Gene Function and Regulation, State Key Laboratory of Biocontrol, School of Life Sciences, Sun Yat-Sen University, Guangzhou, 510275, PR China;
2. RNA Biomedical Institute, Sun Yat-Sen Memorial Hospital, Sun Yat-Sen University, Guangzhou, 510275, PR China; 3. Corresponding author: zhangrui3@mail.sysu.edu.cn

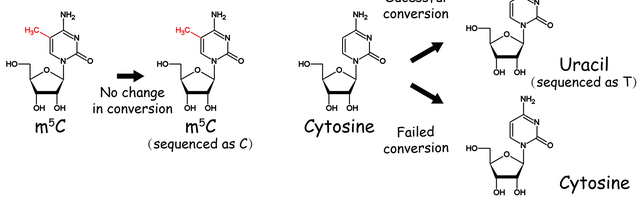
Introduction

1. m^{5}C is one of the longest-studied RNA modifications and is catalyzed by DNMT2 and NSUN methyltransferase family.

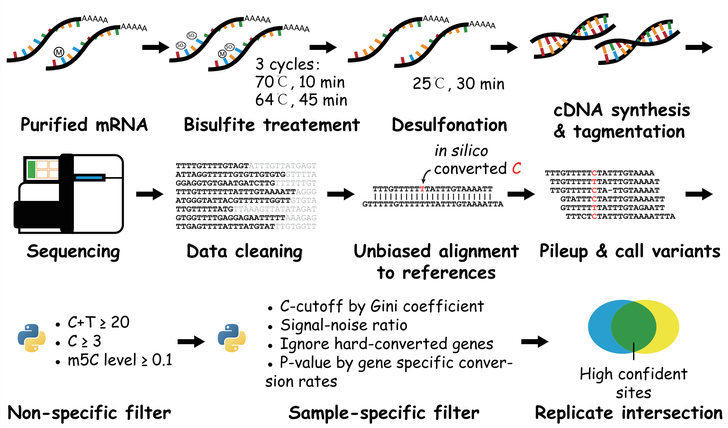
2. RNA bisulfite sequencing (BS-seq) is a method used in m^{5}C methylation research.

3. No common set of mRNA m^{5}C substrate or consensus methylation target sequences were identified by previous studies utilizing RNA BS-seq, indicating that some of the results were artifacts introduced by incomplete bisulfite conversion of structural RNAs or improper analysis of the sequencing data.

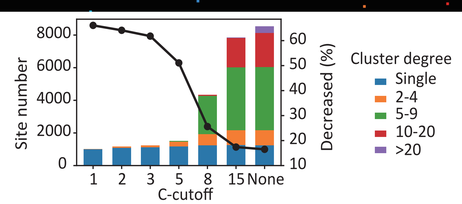
4. A major challenge and demand in mRNA m^{5}C stuides is to develop a robust framework to distinguish *bona fide* mRNA m^{5}C events from noise.



Our improved methodology



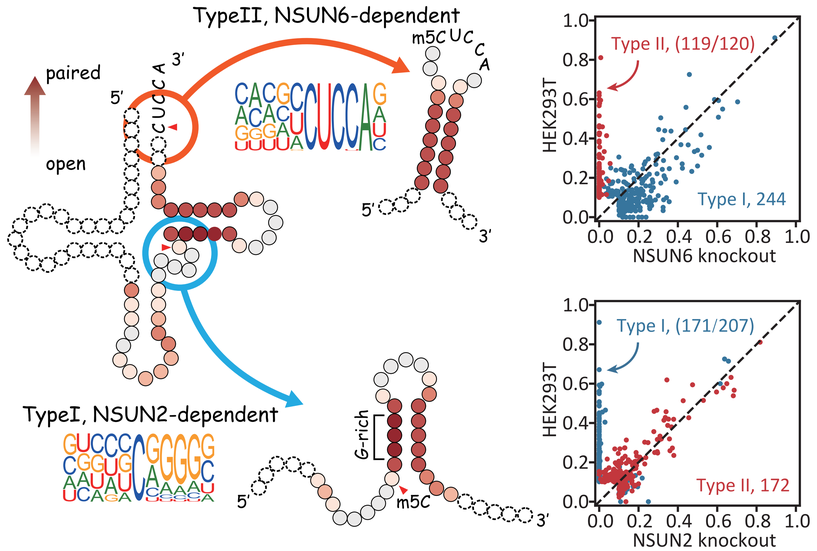
The majority of false positive is in cluster



The relationship between the cluster status of the sites and the proportion of the sites with decreased m^{5}C levels upon NSUN2 knockdown in HeLa cells. Bars: number of sites in different cluster degrees; line, the proportions of the sites with significantly decreased m^{5}C levels. C-cutoff: removing reads with non-converted Cs above a cutoff.

NSUN2 and NSUN6 are major mRNA m^{5}C methyltransferases

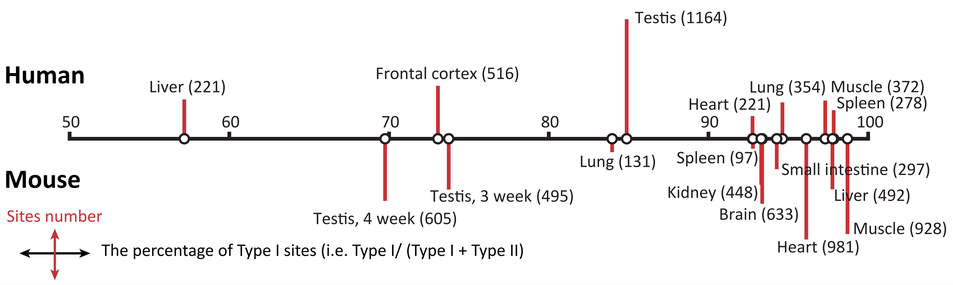
We classified mRNA m^{5}C sites into two groups based on their sequence and structure features: Type I m^{5}C sites are enriched with a 3' G-rich motif adjacent to a small stem loop structure; Type II m^{5}C sites are usually found to be in a 5'-CUCCA-3' motif in a loop.



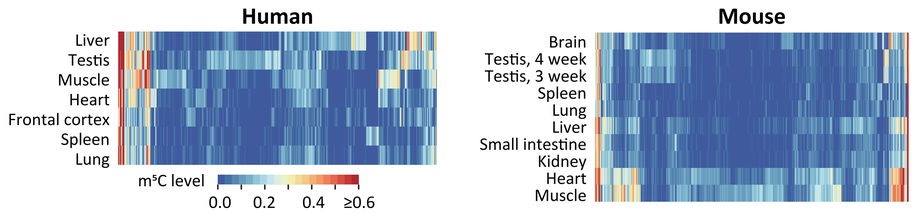
The sequence and structural features of Type I and Type II sites resemble the features recognized by NSUN2 and NSUN6 in tRNAs, respectively.

A comparison of the m^{5}C methylation levels of Type I and Type II sites between wild-type and NSUN2 or NSUN6 knockout HEK293T cells verified that NSUN2 is the writer of Type I m^{5}C and NSUN6 is the writer of Type II m^{5}C.

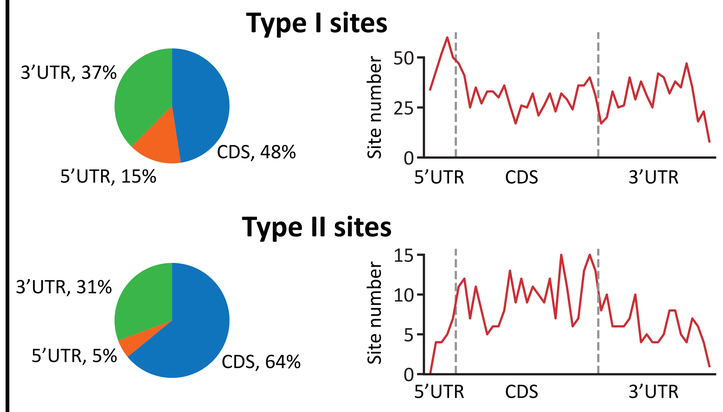
Number and composition of m^{5}C sites in tissues



m^{5}C levels in human and mouse tissues



The distribution of Type I and Type II sites



*human data only, similar distribution in mouse.

Conclusions

- We optimized the experimental and computational workflow for BS-seq to identify *bona fide* m^{5}C sites in mRNA.
- NSUN2 methylates Type I m^{5}C; and NSUN6 methylates Type II m^{5}C.
- The mRNA substrates of NSUN2 and NSUN6 resemble their tRNA substrates in motif and structure.
- In human and mouse tissues, we profiled 97 to 1164 m^{5}C sites. They often have median m^{5}C level around 15%-20%.
- Type I sites enrich around start codon while Type II sites distribute evenly.

References

1. Huang, T., W. Chen, J. Liu, N. Gu and R. Zhang (2019). "Genome-wide identification of mRNA 5-methylcytosine in mammals." Nature Structural & Molecular Biology 26(5): 380-388.

2. Liu, J., T. Huang, Y. Zhang, T. Zhao, X. Zhao, W. Chen and R. Zhang (2020). "Sequence- and structure-selective mRNA m5C methylation by NSUN6 in animals." National Science Review.

3. The pipeline: github.com/SYSU-zhanglab/RNA-m5C

Acknowledgements

This study is funded by the National Key R&D Program of China, Guangdong Major Science and Technology Projects, Guangdong Innovative and Entrepreneurial Research Team Program and National Natural Science Foundation of China.