Genes & Diseases (2023) 10, 1157–1160

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.keaipublishing.com/en/journals/genes-diseases

RAPID COMMUNICATION

Human RNA Modifications Disease Database (HRMDD): A web resource for the molecular and clinical landscape of RNA modifications in human diseases

Here, we developed a comprehensive web-searchable database, designated as Human RNA Modifications Disease Database (HRMDD, http://bio-bigdata.hrbmu.edu.cn/HRMDD/home.jsp). RNA modification (RM) is an important mechanism of epigenetic regulation. With the evolution of both experimental technologies and computational methods, major progress has been made in identifying the genomic locations and distributions of various RM types throughout the transcriptome.[1] Additional breakthroughs came from the identification and characterization of RM regulators. RMs are generally regulated by three different types of regulators, which are deposited, removed, and recognized by proteins known as "writers", "erasers" and "readers", respectively.[2] Increasing evidence suggests that dysregulation of RMs and their regulators is implicated in various cancers, as well as other diseases.[3,4] Despite these advances in understanding RMs and their regulators, their biological functions and mechanisms in human diseases remain largely unknown.

In this work, we systematically collected experimentally supported RM-disease associations. A total of 2082 experimentally supported associations, covering 35 types of RMs, 92 RM regulators, 266 human diseases, 27 virus species and 16 regulatory mechanisms, were manually collected. We herein list the top 10 RM types, diseases, RM regulators and regulatory mechanisms with the largest number of RM-disease associations as hotspot data (Fig. S1A–D). Given the critical functions of RM regulators in cancer, we next assessed the molecular and clinical landscape of RM regulators across 33 cancer types from TCGA project (Fig. S2A, B).

We first assessed the somatic mutations of RM regulators across TCGA cancer types. The overall mutation frequency of RM regulators ranged from 0.1% to 11.89% (Fig. S3). Cancer types such as UCEC, COAD, and STAD, exhibited higher mutation frequencies in RM regulators, whereas PCPG, TGCT, and THYM exhibited lower mutation frequencies (Fig. S2C). Certain regulators, such as *ZC3H13*, *EEF2*, *EIF3A*, *IGF2BP1*, and *YTHDC1/2* (all of which are m^6A regulators) showed higher mutation frequencies (Fig. 1A). We found that readers exhibited relatively higher average mutation frequencies across TCGA cancer types than the writers and erasers (Fig. S2D).

We next assessed the copy number variation (CNV) frequency of RM regulators across TCGA cancer types (Fig. 1B). We found that CNV are prevalent in RM regulators. The RM regulators with the highest (top 10%) and lowest (bottom 10%) CNV frequencies across 33 cancer types are shown in Figure S4A and S4B. RM regulators such as *ADAR* (A-to-I writer), *ALYREF* (m^5C reader), *IGF2BP2* (m^6A reader), and *TRMT12* (yW writer) showed widespread CNV amplifications across cancer types, while *ADAT3* (A-to-I writer), *RBM15* (m^6A writer), *EEF2* (m^6A reader), *ZC3H13* (m^6A writer) and *HENMT1* (Nm writer) showed prevalent CNV deletions across cancer types.

We explored the gene expression alterations of RM regulators across TCGA cancer types. We found that 89% (81/91) of the known regulators were differentially expressed in at least one cancer type, suggesting that the dysregulation of RM regulator expression is prevalent in human cancers (Fig. S5A). Intriguingly, although readers exhibited higher mutation frequencies, writers tended to be dysregulated in more cancer types. The distribution of dysregulated regulators indicated that erasers showed global downregulation across cancer types, whereas writers and readers showed almost equally distributed expression patterns (up/

Peer review under responsibility of Chongqing Medical University.

https://doi.org/10.1016/j.gendis.2022.08.024

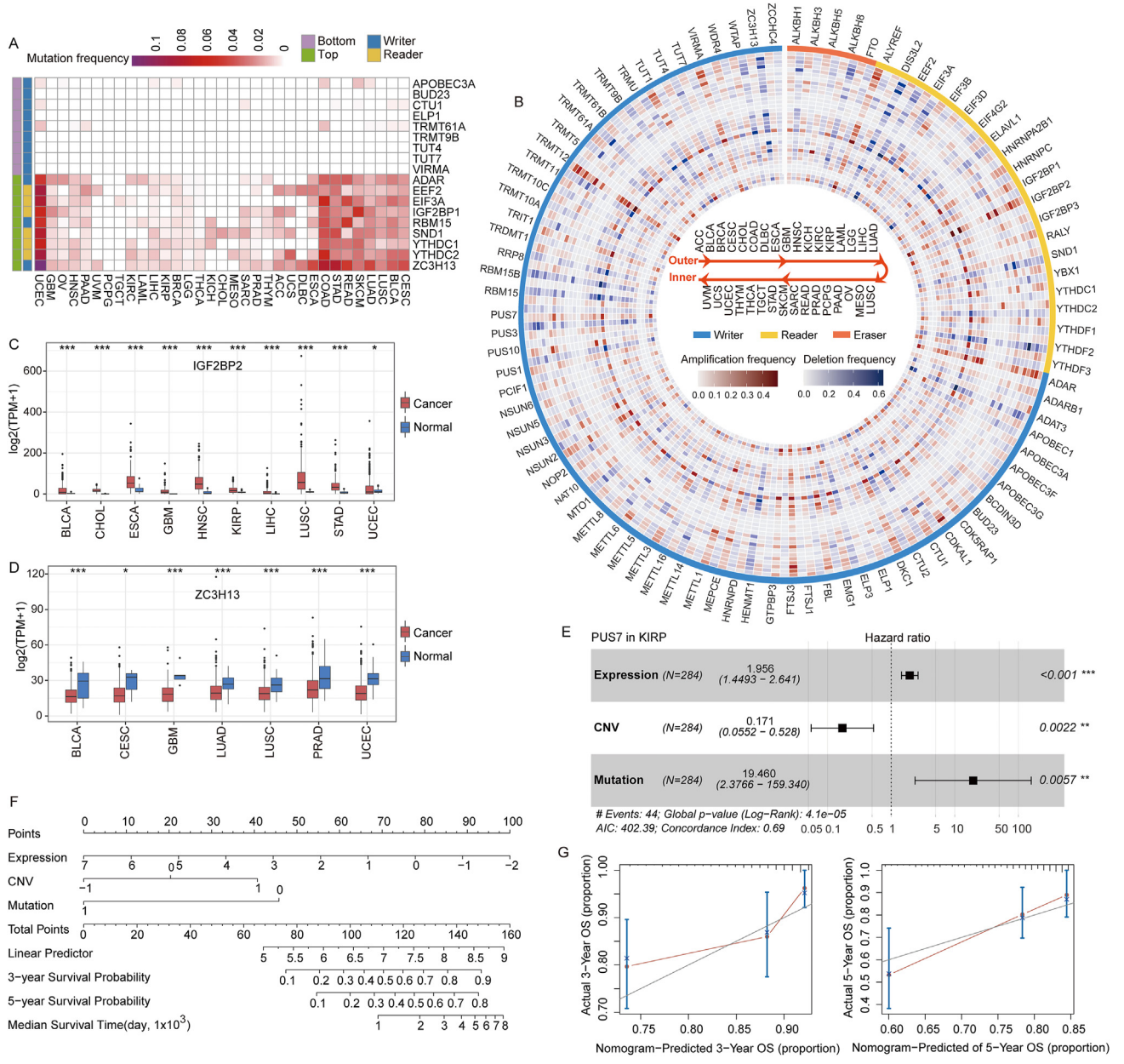

Figure 1 The molecular and clinical relevance of RNA modification regulators across cancer types. **(A)** RM regulators with the highest (top 10%) and lowest (bottom 10%) mutation frequencies across the 33 cancer types in TCGA. **(B)** The CNV alteration frequency of RM regulators across these 33 cancer types. The Circos plot shows the CNV alteration frequency, with red representing the amplification frequency and blue representing the deletion frequency. Each Circo represents one cancer type, which is shown in the inner panel. **(C)** Box plots showing the expression distribution of *IGF2BP2* across tumour (red) and normal samples (blue) in 10 cancer types. **(D)** Box plots showing the expression distribution of *ZC3H13* across tumour (red) and normal samples (blue) in 7 cancer types. *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$. **(E)** A forest plot of the multivariate Cox regression analysis for *PUS7* in KIRP. **(F)** A nomogram composed of the expression, somatic mutation and CNV of *PUS7* for predicting the 3-year and 5-year OS probability in patients with KIRP. **(G)** A calibration plot evaluating the nomogram's prediction of the 3-year and 5-year OS probability based on the *PUS7* status in patients with KIRP.

downregulation) (Fig. S5B). When the distribution of differentially expressed regulators was examined among the different cancer types, we found that READ displayed the largest number of differentially expressed regulators, followed by CHOL, COAD, and LUSC (Fig. S5C).

We further evaluated whether the expression of RM regulators was affected by genetic alterations. Notably, certain RM regulators with relatively higher CNV amplification frequencies were upregulated in tumour samples, while certain RM regulators with relatively higher deletion

frequencies were downregulated in tumour samples. For example, *IGF2BP2*, with higher CNV amplification frequencies in BLCA, HNSC, LUSC, LIHC, STAD, CHOL, GBM, KIRP, ESCA and UCEC, was significantly upregulated in these cancer types (Fig. 1C). In contrast, *ZC3H13*, which had higher CNV deletion frequencies in BLCA, LUSC, LUAD, PRAD, UCEC, GBM and CESC, was significantly downregulated in these cancer types (Fig. 1D). Taken together, these observations suggest that the genetic alterations and differential expression of RM regulators are not only heterogeneous in different cancer contexts, but also have complex regulatory patterns.

Given the important associations of RM regulators with cancer, we explored the clinical relevance of RM regulators across cancer types. We first analysed the associations between the expression of RM regulators and patient survival across the 33 cancer types using the univariate Cox proportional hazards (PH) model. We found that each of the RM regulators affected the patients' overall survival (OS) for at least one type of cancer (Fig. S6). We found that some RM regulator genes were associated with worse survival for patients with several different cancer types. For example, higher expression of *PUS1* (a Psi writer) was associated with worse survival across seven cancer types (Fig. S7). These findings suggest that the RM regulators have prognostic significance and may represent novel prognostic biomarkers.

To improve the prognostic accuracy of the model, we performed a multivariate Cox regression analysis to assess the correlations between RM regulators and patient survival by integrating somatic mutation, CNV, and expression data (Fig. S8). Two regulator genes, *PUS7* and *IGF2BP2*, were found to be correlated with the prognosis of patients in the largest number of cancer types (Fig. 1E; Fig. S9, 10). We identified three additional regulator genes, *ZCCHC4* (m^6A writer), *CDK5RAP1* (ms^2i^6A writer), and *CDKAL1* (ms^2t^6A writer), that showed statistical significance in specific cancer types (Fig. 1E; Fig. S11). Taking *PUS7* in KIRP as an example, the expression and mutation of *PUS7* were found to be risk factors for the OS, whereas CNV was noted to be a protective factor (Fig. 1E). We constructed a nomogram to predict the 3-year and 5-year OS probability, as well as the median survival time by including the mutation, CNV, and expression of *PUS7* in KIRP (Fig. 1F). The calibration plot showed that the predicted OS probability deviated very little from the actual OS probability (Fig. 1G). These results suggest that RM regulators may have potential applications for determining the prognosis of patients with certain cancers, or may represent novel targets for cancer therapy.

To facilitate browsing, searching, downloading, and visualizing data regarding the relationship between RM and human diseases, we developed HRMDD (http://bio-bigdata.hrbmu.edu.cn/HRMDD/home.jsp) (Fig. S12, S13). Regulator-Tool was developed to characterize and visualize the functions of RM regulator genes in cancers based on -omics datasets from TCGA and other resources. Regulator-Tool includes the following functions: (i) the Mutation function enables users to obtain the mutation frequency and proportion of mutation types in an individual regulator in a specific cancer; (ii) the CNV function allows users to obtain the CNV alteration frequency of a regulator in a specific cancer; (iii) the Expression function makes it possible for users to perform differential regulator expression analysis on a specific RM regulator and to generate box plots for cancer and normal groups; (iv) the Correlation function provides a correlation analysis based on regulator expression and scatter diagrams for two correlated genes in a specific cancer; (v) the Survival function performs a Cox regression analysis and produces Kaplan—Meier survival curves for a specific regulator in the selected cancer; (vi) the Network function provides visualization of a regulator-protein network, as well as providing functional annotation; and (vii) the Protein-RNA function provides a list of potential targets of RM regulators of interest.

Author contributions

LW and SN conceived and supervised the project. JW, HZ and XW performed the data analyses. LL, SL, LT, DL and YY developed the database. JW and HZ wrote the paper. All the authors read and approved the final manuscript.

Conflict of interests

The authors have declared no competing interests.

Funding

This work was supported by the National Natural Science Foundation of China (Nos. 81820108014, 82071407, 82171396, and 32070672); National Key Research and Development Project (China) (No. 2018YFE0114400); Heilongjiang Provincial Natural Science Foundation (China) (No. YQ2021H012); The Postdoctoral Foundation of Heilongjiang Province (China) (No. LBH-TZ1019).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.gendis.2022.08.024.

References

1. Helm M, Motorin Y. Detecting RNA modifications in the epitranscriptome: predict and validate. *Nat Rev Genet*. 2017;18(5):275—291.
2. Zaccara S, Ries RJ, Jaffrey SR. Reading, writing and erasing mRNA methylation. *Nat Rev Mol Cell Biol*. 2019;20(10):608—624.
3. Jonkhout N, Tran J, Smith MA, Schonrock N, Mattick JS, Novoa EM. The RNA modification landscape in human disease. *RNA*. 2017;23(12):1754—1769.
4. Barbieri I, Kouzarides T. Role of RNA modifications in cancer. *Nat Rev Cancer*. 2020;20(6):303—322.

Jianjian Wang [a,b,1], Lifang Li [a,1], Hanxiao Zhou [b,1], Shuang Li [a,1], Liting Tian [a,1], Xinyue Wang [b], Danyang Li [a], Yue Yin [a], Shangwei Ning [b,**], Lihua Wang [a,*]

[a] *Department of Neurology, The Second Affiliated Hospital of Harbin Medical University, Harbin, Heilongjiang 150081, China*

[b] *College of Bioinformatics Science and Technology, Harbin Medical University, Harbin, Heilongjiang 150081, China*

*Corresponding author.
**Corresponding author.
E-mail addresses: ningsw@ems.hrbmu.edu.cn (S. Ning), wanglh211@163.com (L. Wang)

8 June 2022
Available online 10 September 2022

[1] Equal contribution.