

RESEARCH

Open Access



# Integrative analyses of the RNA modification machinery reveal tissue- and cancer-specific signatures

Oguzhan Begik<sup>1,2,3</sup>, Morghan C. Lucas<sup>1,4</sup>, Huanle Liu<sup>1</sup>, Jose Miguel Ramirez<sup>1,4</sup>, John S. Mattick<sup>2,3</sup> and Eva Maria Novoa<sup>1,2,3,4\*</sup>

\* Correspondence: [eva.novoa@crg.eu](mailto:eva.novoa@crg.eu)

<sup>1</sup>Center for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, 08003 Barcelona, Spain

<sup>2</sup>Garvan Institute of Medical Research, Darlinghurst, NSW 2010, Australia

Full list of author information is available at the end of the article

## Abstract

**Background:** RNA modifications play central roles in cellular fate and differentiation. However, the machinery responsible for placing, removing, and recognizing more than 170 RNA modifications remains largely uncharacterized and poorly annotated, and we currently lack integrative studies that identify which RNA modification-related proteins (RMPs) may be dysregulated in each cancer type.

**Results:** Here, we perform a comprehensive annotation and evolutionary analysis of human RMPs, as well as an integrative analysis of their expression patterns across 32 tissues, 10 species, and 13,358 paired tumor-normal human samples. Our analysis reveals an unanticipated heterogeneity of RMP expression patterns across mammalian tissues, with a vast proportion of duplicated enzymes displaying testis-specific expression, suggesting a key role for RNA modifications in sperm formation and possibly intergenerational inheritance. We uncover many RMPs that are dysregulated in various types of cancer, and whose expression levels are predictive of cancer progression. Surprisingly, we find that several commonly studied RNA modification enzymes such as METTL3 or FTO are not significantly upregulated in most cancer types, whereas several less-characterized RMPs, such as LAGE3 and HENMT1, are dysregulated in many cancers.

**Conclusions:** Our analyses reveal an unanticipated heterogeneity in the expression patterns of RMPs across mammalian tissues and uncover a large proportion of dysregulated RMPs in multiple cancer types. We provide novel targets for future cancer research studies targeting the human epitranscriptome, as well as foundations to understand cell type-specific behaviors that are orchestrated by RNA modifications.

**Keywords:** RNA modifications, Epitranscriptomics, Tissue specificity, Dysregulation in cancer

## Background

Technological advancements have revolutionized our understanding of RNA modifications, which can occur by removal (by deamination, often called “RNA editing”) or by



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

the addition of chemical side groups on the ribose or base moieties. These chemical entities, collectively known as the “epitranscriptome” [1], not only occur in tRNAs and rRNAs, where they were first identified and have traditionally been studied, but also in other molecules, such as mRNAs, long noncoding RNAs, piRNAs, and miRNAs [2–6]. A number of studies have shown that RNA modifications can profoundly affect central biological processes, including cell fate [7], sex determination [8, 9], maternal-to-zygotic transition [10], and the circadian clock [11] as well as plant developmental timing, morphogenesis, and flowering [12]. Furthermore, dysregulation of their activity has been associated with more than 100 different human diseases [13–17]. At a molecular level, modifications can affect the fate and function of the RNA molecules that contain them, including turnover rates [18–20], translation efficiency [21, 22], and subcellular localization [23], among others.

Over 170 different RNA modifications are known to decorate RNA molecules [24]. In the last few years, a vast amount of efforts have been devoted to functionally dissecting the biological role of N6-methyladenosine ( $m^6A$ ), the most prevalent internal RNA modification found in human mRNAs.  $M^6A$  is placed by a multicomponent transferase complex, in which methyltransferase 3 (METTL3) acts as the catalytic subunit [25, 26]. Moreover,  $m^6A$  modifications can be reversed by the activity of  $m^6A$  demethylases or “erasers,” namely the fat mass and obesity-associated protein (FTO) [27] and the alkB homolog 5 (ALKBH5), although recent studies have suggested that only the latter can demethylate  $m^6A$  marks [28]. Mechanistically,  $m^6A$  modifications can alter mRNA splicing [29–31], cause mRNA decay [20], and affect translation [2, 32, 33]; thus, they can govern major cellular processes including cellular fate [34, 35], stress responses [2], and differentiation programs [36]. These features have set out  $m^6A$  marks, and more specifically, their “readers,” “writers,” and “erasers,” as promising drug targets for multiple diseases, including cancer [37–40]. However, the functional characterization of the majority of RNA modifications still remains an uncharted territory.

Insights into the physiological roles of specific RNA modification-related proteins (RMPs) have mostly come from naturally occurring phenotypes or diseases associated with their loss of function [13–17]. However, a systematic annotation and characterization of RNA modification RMPs across human tissues, cell types, and disease states is currently lacking.

Here, we have compiled and analyzed the evolutionary history of 90 RNA modification writers as well as the gene expression patterns of 146 human RMPs (Additional file 1: Table S1) from 32 tissues, 10 species, and 13,358 tumor-normal samples. Our analyses revealed that many RMPs display restricted gene expression patterns and/or are dysregulated in specific types of cancer. Specifically, we found that a vast proportion of RNA modification “writers” have undergone duplications (84%) and that these were typically accompanied by a change in their RNA target specificity and/or tissue expression patterns (82%). We observed that the most frequent change in tissue specificity is the acquisition of restricted testis-specific expression, suggesting that a significant portion of the human RNA modification machinery is likely devoted to sperm formation and maturation. We also found that 27% of human RMPs are significantly dysregulated in cancers, and identified several dysregulated RMPs whose expression is strongly correlated with cancer prognosis. Overall, our work reveals an unanticipated heterogeneity of RMP expression across both normal and malignant cell types, and

points towards several less-characterized RMPs, such as HENMT1 or LAGE3, as promising drug targets for antitumor therapies.

## Results

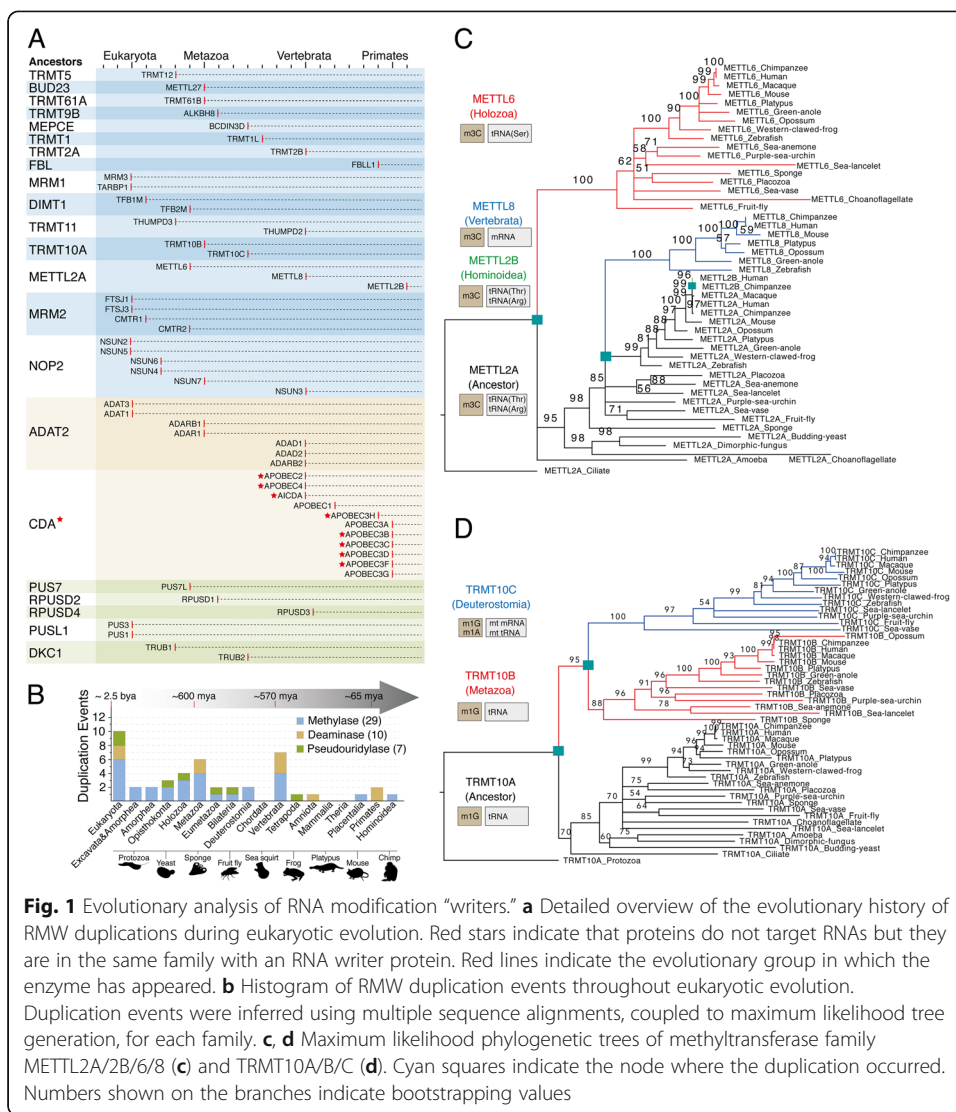
### Comprehensive annotation and evolutionary analysis of RNA modification writers

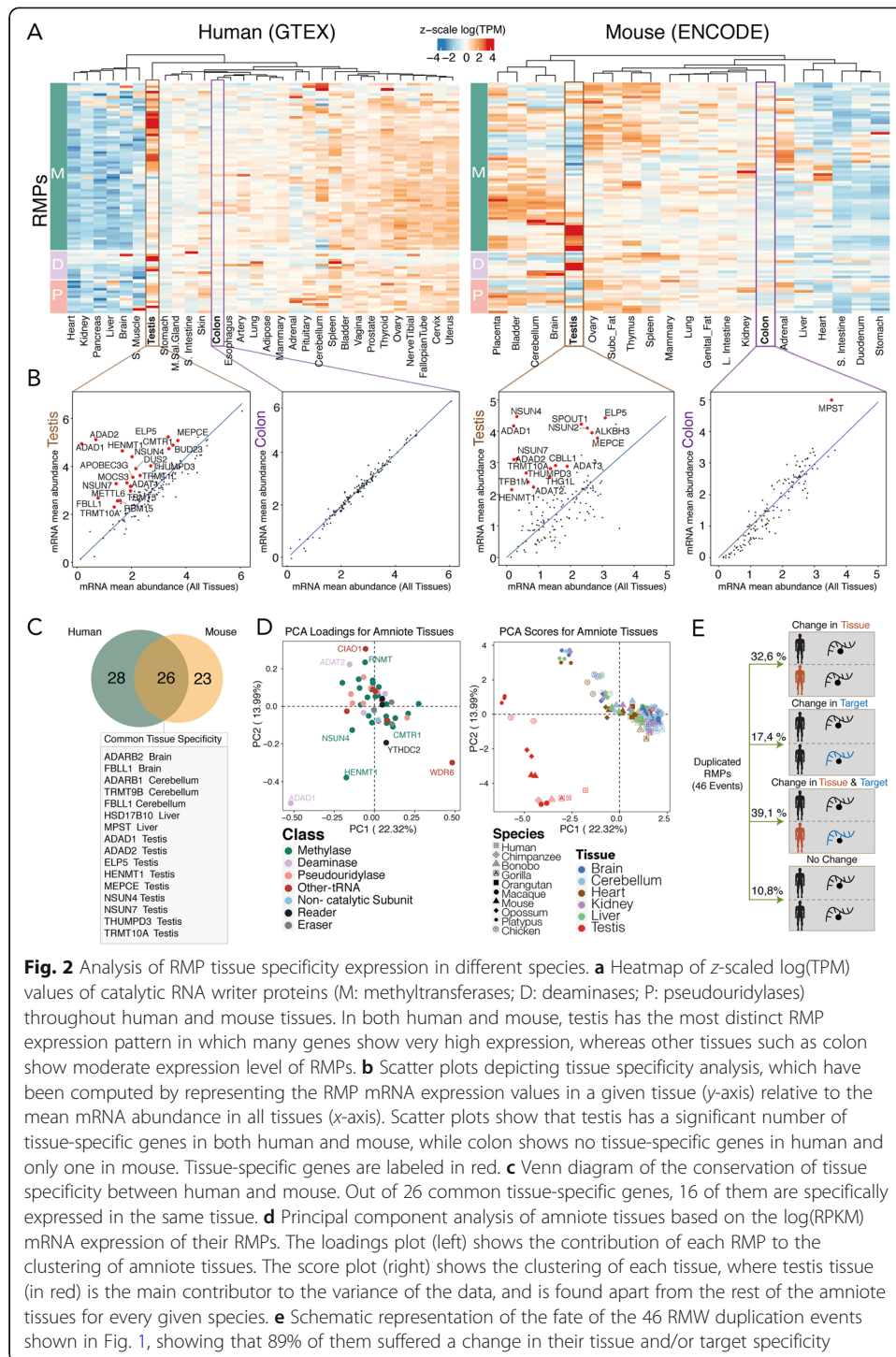
To reveal the evolutionary history of the RNA modification machinery, we first compiled and manually curated a list of human RMPs (Additional file 1: Table S1, see also “Methods”). Due to the wide chemical variety of RNA modifications, we restricted our evolutionary analysis to the catalytic domain of three major RNA modification “writer” (RMW) classes: (i) methyltransferases, (ii) pseudouridylases, and (iii) deaminases. For each annotated RMW [13, 41, 42], Pfam domains of the catalytic domain were extracted and used as input for HMM-based searches against the human proteome. This resulted in a total of 90 human RMWs, doubling the amount of annotated human RMWs in other resources [41]. To determine the evolutionary history and identify duplication events that occurred in each family, ortholog proteins from representative species were retrieved (see “Methods”), and phylogenetic trees were built to identify the number of duplications occurring within each family. Overall, our analysis identified 46 duplication events (Fig. 1a), which have mainly occurred in the base of Eukaryota, Metazoa, and Vertebrata (Fig. 1b, see also Additional file 2: Table S2).

We find that duplications are often accompanied by changes in substrate specificity (Fig. 1c, d), at least in those RMWs where the substrate specificity has been reported. One such case is the family of 3-methylcytosine ( $m^3C$ ) RNA methyltransferases, where the ancestral protein methyltransferase-like protein 2 (METTL2) modifies both  $tRNA^{Arg}$  and  $tRNA^{Thr}$ , whereas its paralog enzymes, METTL6 and METTL8, methylate  $tRNA^{Ser}$  and mRNA, respectively [43] (Fig. 1c). Similarly, we find that the N1-methylguanosine ( $m^1G$ ) methyltransferases TRMT10A and TRMT10B modify tRNAs in position  $m^1G9$  [44], whereas its paralog TRMT10C has been reported to place N1-methyladenosine ( $m^1A$ ) in mitochondrial tRNAs and mRNAs [3], in addition to  $m^1G$  in tRNAs (Fig. 1d).

### Heterogeneity of expression patterns among duplicated RMPs is conserved across species

We then wondered whether duplicated RMPs might have acquired distinct tissue expression patterns than the ancestral gene. To test this, we examined the heterogeneity of RMP expression patterns across tissues in human and mice, using publicly available RNASeq datasets [45–47] (Fig. 2a, see also Additional file 13: Figure S1A for gene-labeled heatmaps). For each gene and tissue, we computed “tissue specificity (TS) scores” [48], which is defined as the deviation of gene expression levels in a given tissue, relative to the average expression across all tissues (see “Methods”). Using this approach, we found that testis is the most distinctive tissue in terms of RMP gene expression patterns, both in human and mouse (Fig. 2a, b). This was due to several RMPs being quasi-exclusively expressed in testis (e.g., ADAD1, ADAD2), but also to several RMPs whose expression levels are significantly increased in this tissue (e.g., FBLL1, HENMT1, NSUN7). In contrast, other tissues such as colon displayed none or few tissue-enriched RMPs (Fig. 2b, see also





**Fig. 2** Analysis of RMP tissue specificity expression in different species. **a** Heatmap of z-scaled log(TPM) values of catalytic RNA writer proteins (M: methyltransferases; D: deaminases; P: pseudouridylases) throughout human and mouse tissues. In both human and mouse, testis has the most distinct RMP expression pattern in which many genes show very high expression, whereas other tissues such as colon show moderate expression level of RMPs. **b** Scatter plots depicting tissue specificity analysis, which have been computed by representing the RMP mRNA expression values in a given tissue (y-axis) relative to the mean mRNA abundance in all tissues (x-axis). Scatter plots show that testis has a significant number of tissue-specific genes in both human and mouse, while colon shows no tissue-specific genes in human and only one in mouse. Tissue-specific genes are labeled in red. **c** Venn diagram of the conservation of tissue specificity between human and mouse. Out of 26 common tissue-specific genes, 16 of them are specifically expressed in the same tissue. **d** Principal component analysis of amniote tissues based on the log(RPKM) mRNA expression of their RMPs. The loadings plot (left) shows the contribution of each RMP to the clustering of amniote tissues. The score plot (right) shows the clustering of each tissue, where testis tissue (in red) is the main contributor to the variance of the data, and is found apart from the rest of the amniote tissues for every given species. **e** Schematic representation of the fate of the 46 RMP duplication events shown in Fig. 1, showing that 89% of them suffered a change in their tissue and/or target specificity

(Fig. 2d, see also Additional file 13: Figure S4). Overall, we found that 89% of RMP duplication events were often followed by a change in tissue specificity (32.6%), target specificity (17.4%), or both (39.1%) (Fig. 2e, see also Additional file 4: Table S4 and Additional file 13: Figure S5), with a major over-representation of acquisition of testis-specific gene expression.

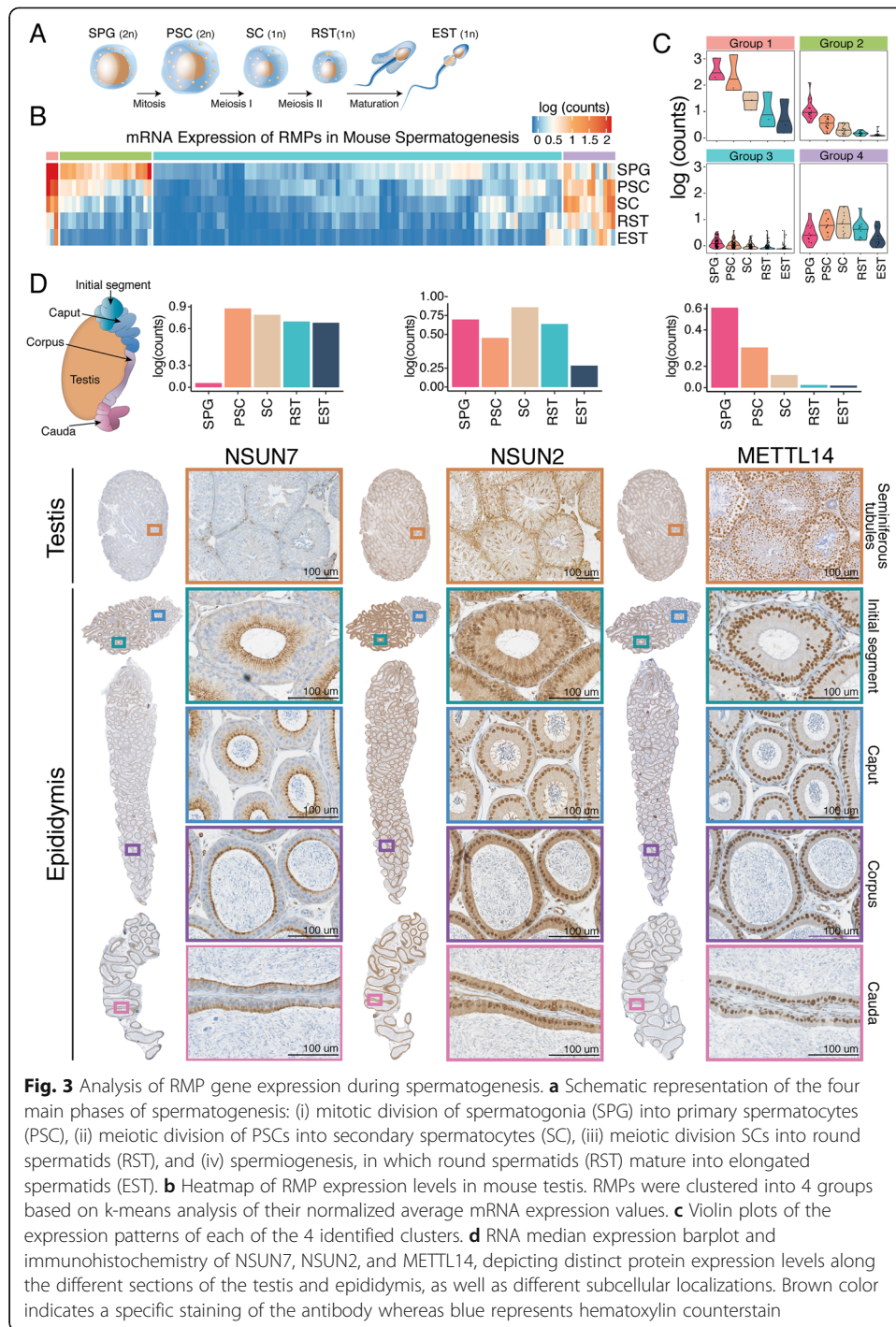
### Testis-specific RMPs are mainly expressed during meiotic stages of spermatogenesis

The process of sperm formation, termed spermatogenesis (Fig. 3a), is a highly specialized differentiation process in which transcriptional, post-transcriptional, and translational regulation are highly orchestrated [50–53]. RNA modifications can influence pre-mRNA splicing, mRNA export, turnover, and translation, which are controlled in the male germline to ensure coordinated gene expression [35]. Recent works have shown that m<sup>6</sup>A depletion in mice dysregulates translation of transcripts that are required for spermatogonial proliferation and differentiation [34]. Moreover, m<sup>5</sup>C modifications have been shown to be essential for transmission of diet-induced epigenetic information across generations in the epididymis [54]. However, whether additional RNA modifications may be involved in such orchestration is largely unknown.

To identify at which stage of sperm formation and maturation testis-specific RMPs are involved in, we gathered publicly available single-cell RNA sequencing data from mouse testis [55] (Fig. 3b, see also Additional file 13: Figure S6A for gene-labeled heatmap). We first classified RMPs based on their gene expression patterns (see “Methods”), identifying four main expression patterns: (i) high expression only during mitotic stages (spermatogonia), (ii) high expression in both mitotic and meiotic stages (spermatogonia, spermatocytes, spermatids), although decreased in the latter, (iii) low expression throughout spermatogenesis, and (iv) high expression only during meiotic stages (spermatocytes and spermatids) (Fig. 3b, c, see also Additional file 13: Figure S6B,C).

We find that the majority of RMPs, including those involved in placing, reading, and removing m<sup>6</sup>A (VIRMA, YTHDC2, YTHDF2, ALKBH5, METTL14, METTL3) are highly expressed in spermatogonial cells, whereas their expression rapidly drops as the spermatogenic process begins (Fig. 3b, c, see also Additional file 13: Figure S6C). Interestingly, this is not the case for all RMPs, such as m<sup>5</sup>C methyltransferase NSUN7, which is not expressed in early stages of spermatogenesis, but whose expression levels are drastically increased in spermatocytes and spermatids (Additional file 13: Figure S6A,C). Similarly, the testis-specific adenosine deaminase ADAD1 is not expressed in early stages of spermatogenesis, but its expression levels are greatly increased in meiotic stages. Depletion of NSUN7 or ADAD1 are known to cause infertility [56, 57], suggesting that RMPs that are selectively expressed in meiotic stages of spermatogenesis are essential for proper sperm formation and/or maturation. However, the molecular mechanisms behind these infertility phenotypes are largely uncharacterized. Similar expression patterns were observed when analyzing other publicly available single-cell mouse spermatogenesis RNAseq datasets [58, 59] (Additional file 13: Figure S7, S8 and S9).

We then investigated whether specific RMPs also showed increased expression patterns in epididymis, relative to other tissues (Additional file 13: Figure S6D). Our analysis identified two RMPs as epididymis-enriched: (i) TRDMT1—also known as DNMT2—a 5-methylcytosine (m<sup>5</sup>C) methyltransferase modifying position 38 in specific tRNAs [60], and (ii) METTL1, a N7-methylguanosine (m<sup>7</sup>G) tRNA methyltransferase, which has been recently shown to act not only on tRNAs, but also on miRNAs, promoting their maturation [61]. Interestingly, TRDMT1 has been shown to play a major role in the transmission of paternal epigenetic information across generations [54]; however, whether METTL1 is involved in the transmission of such information is yet to be determined.



### Immunohistochemistry reveals heterogeneity in RMP expression patterns along the epididymis

It is well established that mRNA levels do not always correlate well with protein levels [62]. Thus, to assess whether our findings would hold at the protein level, we performed immunohistochemistry in both testis and epididymis to characterize the expression patterns of 4 RMPs at the protein level: (i) NSUN7, a putative m<sup>5</sup>C methyltransferase that has been shown to affect sperm motility [57, 63]; (ii) NSUN2, an

m<sup>5</sup>C tRNA methyltransferase involved in sperm differentiation [64]; (iii) METTL14, a component of the m<sup>6</sup>A methyltransferase complex, which has been shown to be dynamically regulated during spermatogenesis [34]; and (iv) HENMT1, a piRNA 2'-O-methyltransferase responsible for transposon silencing during spermatogenesis [5] (Fig. 3d, see also Additional file 13: Figure S6E).

We found that NSUN7 is most highly expressed in spermatocytes, as well as in the initial segment and caput regions of the epididymis, in agreement with its role in the acquisition of sperm motility [57, 63, 65, 66] (Fig. 3d, left panels). Intriguingly, NSUN7 displayed vesicular-like localization in the epithelial cells of epididymal ducts, with significant accumulation in the apical surface. It is yet to be determined how NSUN7 depletion causes defects in sperm motility, as well as which are the targets of NSUN7 in testis and epididymal tissues. On the other hand, NSUN2 displayed high expression levels in spermatocytes and spermatids (Fig. 3d, middle panels). We also observed that NSUN2 is highly expressed in the initial segment of the epididymis, with decreased expression in the remaining epididymal sections. To identify the subcellular localization of NSUN7 and NSUN2, we performed immunofluorescence assays in mice testis, co-staining with either fibrillarin (FBL, nucleolar marker) or DDX4 (chromatoid body marker [67]) (Additional file 13: Figure S29). We observed that NSUN2 was mainly expressed in the adluminal compartment in later stages of spermatogenic maturation in seminiferous tubules. Surprisingly, we found that the expression of NSUN2 and DDX4 was quasi mutually exclusive, being DDX4 expressed in earlier stages of spermatogenesis, and NSUN2 being expressed in later stages. We did not observe colocalization of NSUN2 with DDX4, in contrast to previous reports [64].

METTL14 was also found to be highly expressed in early spermatogenesis and down-regulated during the subsequent stages at the mRNA level (Fig. 3d, right panels), in agreement with the dynamic regulation of m<sup>6</sup>A levels during spermatogenesis [34]. This result was corroborated at the protein level using IHC, where METTL14-positive early spermatogenic cells are found in the periphery of the seminiferous tubules, while round spermatids and elongated spermatids, located in the very interior of the seminiferous tubules, and spermatozoa, found in the lumen of the seminiferous tubules and epididymis (see Additional file 13: Figure S10), were negative. Finally, HENMT1 was mostly highly expressed in spermatogonia and secondary spermatocytes at the RNA level; however, IHC of HENMT1 did not show stage- or cell-specific staining (Additional file 13: Figure S6E). Overall, our analyses showed that RMPs are dynamically expressed during spermatogenesis and during sperm maturation and that, for the four genes investigated, protein expression patterns were largely in agreement with mRNA expression.

#### **Analysis of RMP expression in tumor-normal paired human samples reveals heterogeneity in RMP dysregulation across cancer types**

Due to their ability to modulate RNA metabolism and influence protein synthesis rates, RNA modifications have recently emerged as important regulators of cancer [68–70]. Several studies have shown that modulation of the RNA modification machinery can decrease the expression of specific oncogenes [36, 71]. For example, in the case of glioblastoma, treatment with an FTO inhibitor was shown to decrease the expression levels



of certain oncogenes [69]. Similarly, tRNA-modifying enzymes NSUN2 and METTL1 can affect chemotherapy sensitivity by changing the methylation states of certain tRNAs [72]. Thus, understanding which epitranscriptomic players are dysregulated in each tumor type is essential to guide the research for future anticancer therapies targeting this regulatory layer.

To this end, we performed an integrative analysis of RMP gene expression across 13,358 tumor-normal paired human samples gathered from publicly available datasets [73], which included 28 different cancer types (Additional file 5: Table S5). Firstly, we compared the expression patterns between paired tumor-normal samples by measuring the log<sub>2</sub> fold changes of median gene expression between tumor and normal paired samples, for each RMP and cancer type (Additional file 13: Figure S11 and “Methods”). We found that certain cancer types, such as pancreatic adenocarcinoma (PAAD) and acute myeloid leukemia (LAML), showed significant dysregulation of a vast proportion of RMPs (Additional file 13: Figure S11). Surprised by this result, we wondered whether these global up-/downregulation patterns could in fact be an artifact generated by the use of external datasets. Indeed, certain TCGA cancer types do not have real “matched” tumor-normal data readily available and often employ data from other publicly available datasets (e.g., GTEx) as “normal” human tissue (Additional File 5: Table S5).

To address this issue, we extracted the gene expression levels of all genes—not just RMPs—for each cancer type, finding that certain cancer types that employ GTEx data as source of “normal” human tissues, such as LAML, display low Pearson correlation values between matched tumor-normal samples ( $r^2 = 0.86$ ), compared to those observed in other cancer types such as prostate adenocarcinoma (PRAD) ( $r^2 = 0.98$ ) (Additional file 13: Figure S12). Thus, to identify which RMPs were significantly dysregulated in each cancer type, we computed “dysregulation scores” [48], which take into account the global variance of the tumor-normal paired data, for each cancer type (Fig. 4a). We considered an RMP as dysregulated in a given cancer type if its dysregulation score was higher than 2.5 standard deviations (SD) relative to the linear fit to the gene expression in the matched normal tissue (see “Methods”). Using this strategy, we identified a total of 40 RMPs which are dysregulated in at least one cancer type (Table 1, see also Additional file 6: Table S6). Moreover, we find that the “global” up-/downregulation patterns found using log<sub>2</sub> fold change comparisons are not further observed (Fig. 4b), suggesting that these results were in fact artifacts caused by the lack of proper “matched” normal tissues for certain cancer types.

#### **Dysregulation score analyses of tumor-normal paired human samples identify LAGE3 and HENMT1 as top-ranked dysregulated RMPs**

We then asked whether specific RMP genes were recurrently up- or downregulated in multiple cancer types, as these could constitute promising drug targets that could be used to treat diverse cancer types. We identified 11 RMPs that were upregulated in two or more cancer types, as well as 8 RMPs which were consistently downregulated in at least 2 cancer types (Fig. 4c, see also Table 1). We found that the most frequently upregulated RMP was HENMT1 (Fig. 4d), a piRNA 2'-O-methyltransferase which is highly expressed in gonadal cells, involved in transposable element (TE) mutagenesis

protection [5, 74, 75]. Whether the global upregulation of HENMT1 in cancer samples might be contributing to increased TE mutagenesis is currently unknown.

The second most frequently upregulated RMP was the L antigen family member 3 (LAGE3), a component of complex responsible for formation of N6-threonylcarbamoyladenine ( $t^6A$ ) in position 37 of tRNAs (Fig. 4d). Interestingly, this modification is found in the anticodon stem-loop of many tRNAs decoding ANN codons [76] and has been shown to affect both translation accuracy as well as efficiency [77]. Upregulation of HENMT1 and LAGE3 expression levels was consistently observed in tumors from distinct stages, with the highest expression in stages III and IV (Additional file 13: Figure S13).

We then examined whether LAGE3 and HENMT1 would be upregulated in patient-derived samples at the protein level. To this end, we employed tissue microarrays (TMAs) in combination with immunohistochemistry, analyzing a total of 72 samples (cores) from both tumor and normal tissues, for 12 different cancer types. Our results show that both LAGE3 and HENMT1 are upregulated in specific tumor types at the protein level (Fig. 5a, b), although the observed differences were not found to be statistically significant (Additional file 13: Figure S14). Nevertheless, our results suggest that LAGE3 and HENMT1 have altered expression levels in specific cancer types also at the protein level.

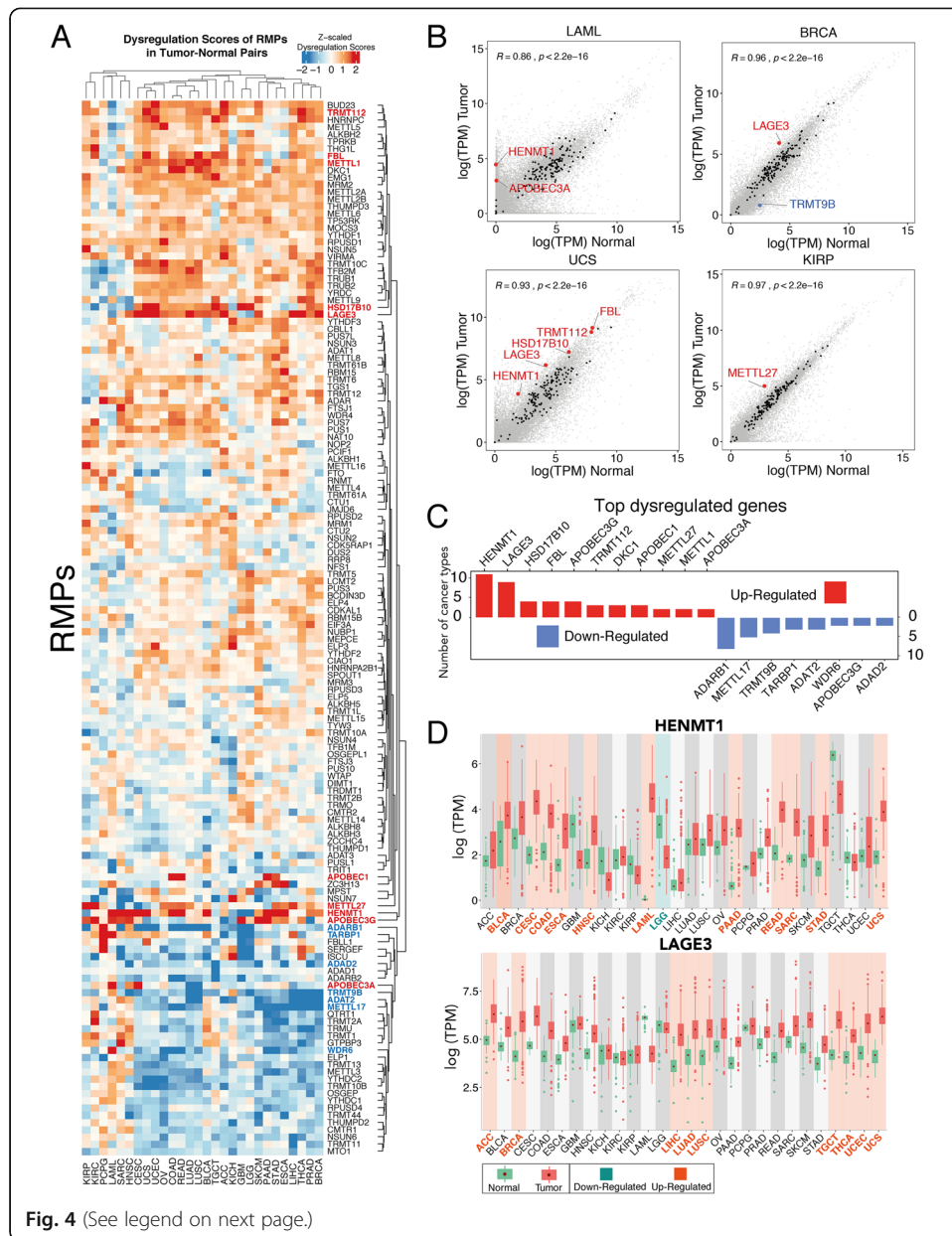
Finally, we asked whether the expression levels of RMPs might be correlated with cancer prognosis. We identified 283 cases where RMP expression patterns are significantly associated with patient survival (Fig. 5c, see also Additional file 7: Table S7). For example, we found that high NSUN5 expression levels in glioblastoma (GBM) are correlated with poor prognosis, in agreement with a recent study [78]. Similarly, our work revealed BUD23 expression to be correlated with cancer survival, in agreement with another recent study [79].

Surprisingly, we found that FTO expression levels are not significantly correlated with patient survival in LAML (Additional file 7: Table S7), despite this cancer type being used to test FTO inhibitors [80]. By contrast, LAGE3 expression levels were significantly correlated with patient survival in LAML (Fig. 5d). Among all the RMP-cancer pairs studied, we identified NSUN7 as the top-ranked RMP in terms of prediction of lower-grade glioma (LGG) patient survival ( $p = 8e^{-24}$ ), although its biological role still remains uncharacterized. Future research will be needed to functionally dissect the role that NSUN7 plays in glioma, as well as to decipher why its expression levels are highly predictive of patient survival.

## Discussion

Over the past decade, systematic efforts to detect and map RNA modifications have boosted the new field of epitranscriptomic research. Many proteins are involved in the writing, reading, and erasing of RNA modifications, but their roles in tumorigenesis and potential as therapeutic targets remain largely uncharacterized. To bridge this gap, here we have compiled a list of 146 human RNA modification-related proteins (RMPs) (Additional file 1: Table S1) and have analyzed the evolutionary history and gene expression patterns of 90 RMPs across 32 mammalian tissues, 10 species, 5 cell types, and 13,358 tumor-normal paired cancer samples.

Through this analysis, we identify a large amount of duplication events in multiple RNA modification families (Fig. 1) and find that duplications are often accompanied by



(See figure on previous page.)

**Fig. 4** Expression analysis of RMPs in human tumor-normal paired samples. **a** Heatmap of z-scaled dysregulation scores of RMPs in tumor-normal paired samples, across 28 cancer types. Positive (red) values indicate upregulation in tumor samples, whereas negative (blue) values indicate downregulation. Genes labeled as red (upregulated) and blue (downregulated) represent top significantly dysregulated genes, which are also individually listed in panel **c**. **b** Scatter plot comparing RMP expression levels of matched tumor-normal samples, for the following cancer types: LAML (acute myeloid leukemia) and UCS (uterine carcinosarcoma) BRCA (breast invasive carcinoma) and KIRP (kidney renal papillary cell carcinoma). Values represent median log(TPM) across all patients. Black data points indicate the expression of RMPs, where dysregulated genes are highlighted in red (upregulated) or blue (downregulated). Non-RMP genes are depicted in gray. **c** Barplot illustrates the number of cancer types in which significantly dysregulated genes are highlighted in red (upregulated) or blue (downregulated). Only RMPs that are dysregulated in more than 2 cancer types are shown. For the full list of dysregulated RMPs, see Table 1. **d** Boxplots of log(TPM) mRNA expression values of HENMT1 (upper panel) and LAGE3 (bottom panel) across all 28 cancer types analyzed in this work. Green box plots represent normal samples, whereas red box plots represent tumor samples. Tumor-normal pairs highlighted in cyan represent cancer types in which the RMP is significantly downregulated, whereas those highlighted in orange represent those cancer types in which the RMP is upregulated. Error bars represent standard deviation of mRNA expression levels across patients. Each data point represents a different patient sample. Abbreviations: ACC (adrenocortical carcinoma), BLCA (bladder urothelial carcinoma), BRCA (breast invasive carcinoma), CESC (cervical squamous cell carcinoma and endocervical adenocarcinoma), COAD (colon adenocarcinoma), ESCA (esophageal carcinoma), GBM (glioblastoma multiforme), HNSC (head and neck squamous cell carcinoma), KICH (kidney chromophobe), KIRC (kidney renal clear cell carcinoma), KIRP (kidney renal papillary cell carcinoma), LAML (acute myeloid leukemia), LGG (brain lower-grade glioma), LIHC (liver hepatocellular carcinoma), LUAD (lung adenocarcinoma), LUSC (lung squamous cell carcinoma), OV (ovarian serous cystadenocarcinoma), PAAD (pancreatic adenocarcinoma), PCCPG (pheochromocytoma and paraganglioma), PRAD (prostate adenocarcinoma), READ (rectum adenocarcinoma), SARC (sarcoma), SKCM (skin cutaneous melanoma), STAD (stomach adenocarcinoma), TGCT (testicular germ cell tumors), THCA (thyroid carcinoma), UCEC (uterine corpus endometrial carcinoma), UCS (uterine carcinosarcoma)

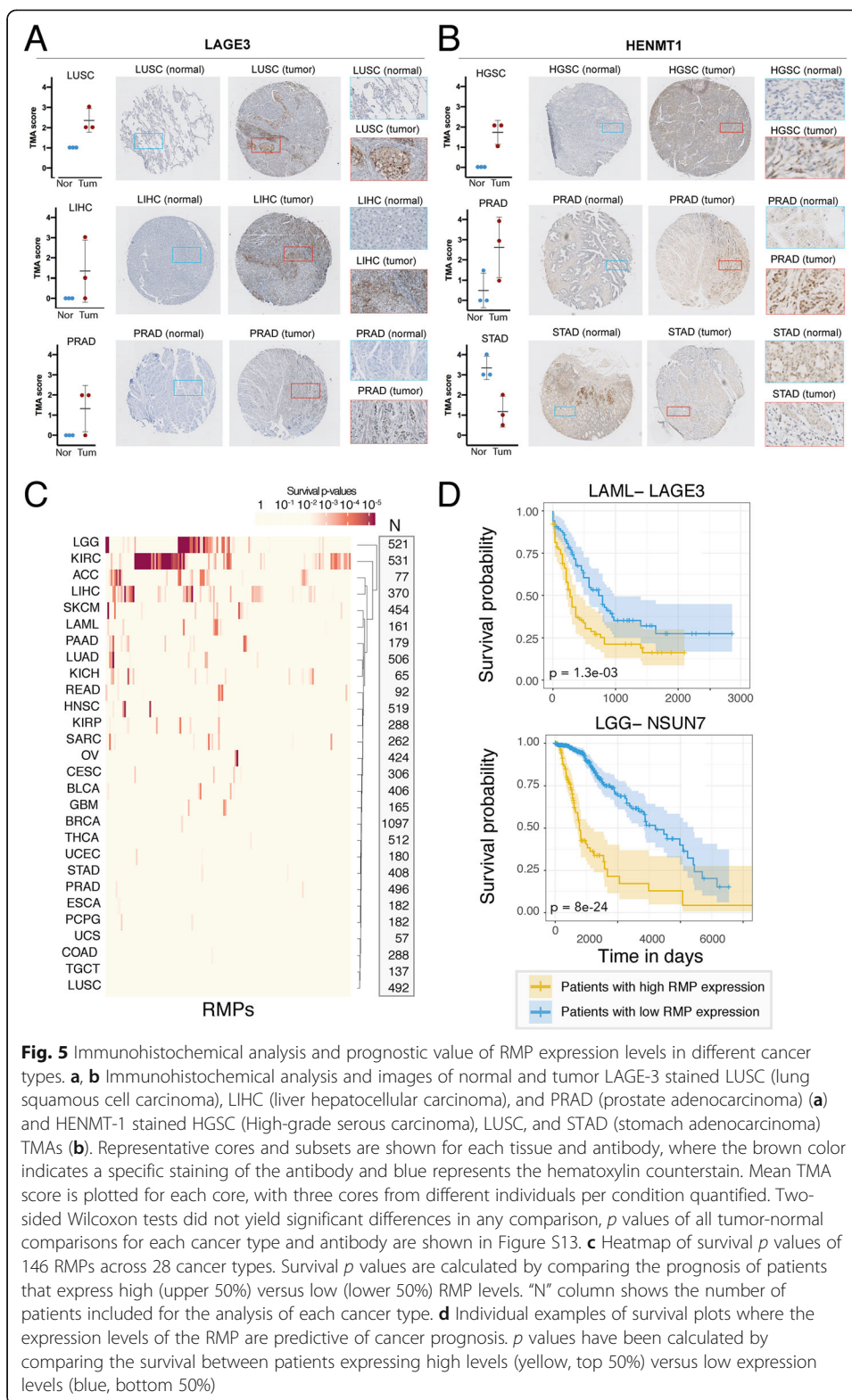
the acquisition of restricted tissue expression patterns and/or change in its RNA target specificity. Therefore, RMP gene duplication is a strategy to acquire novel functions and is typically achieved by altering the expression patterns and/or RNA target selectivity of the paralog proteins, in agreement with other works studying gene evolution [81]. We find that the majority of tissue-restricted RMPs are in fact testis-enriched (Fig. 2), suggesting that certain RMPs might play a pivotal role in sperm formation and maturation. Moreover, deletion of testis-enriched genes such as NSUN7, ADAD1, or HENMT1 leads to male sterility [5, 56, 57, 63].

At the beginning of spermatid elongation, nuclear condensation starts, and consequently, the transcriptional machinery is shut down. Therefore, to provide proteins for the following maturation steps of sperm assembly, mRNAs have to be premade in spermatocytes and round spermatids, before nuclear condensation happens, and translationally repressed until needed [50–53]. Chemical RNA modifications provide an ideal platform to achieve the fine regulation that is required upon transcriptional shutdown, determining which RNAs are expressed, repressed, or undergo decay [82]. In this regard, previous work has shown that METTL3/METTL14 mediated m<sup>6</sup>A modification is dynamically regulated in spermatogenesis [34]. Similarly, piRNA molecules in germline cells are tightly regulated by HENMT1, via 2'-O-methylation of their 3' ends [5]. Here we show that a vast proportion of RMPs are dynamically regulated during spermatogenesis as well as during sperm maturation in the epididymis and, as such, may be involved in the regulation and decay of specific transcripts that occur during sperm formation and maturation (Fig. 3).

**Table 1** List of significantly dysregulated RMPs identified using dysregulation score-based analysis

Short name	Cancer type	Upregulated RMPs	Downregulated RMPs
ACC	Adrenocortical carcinoma	BUD23, FBL, LAGE3, TRMT112, VIRMA	ADAT2, APOBEC3G, TARBP1
BLCA	Bladder urothelial carcinoma	HENMT1	ADARB1
BRCA	Breast invasive carcinoma	LAGE3	TRMT9B
CESC	Cervical squamous cell carcinoma and endocervical adenocarcinoma	APOBEC3A, HENMT1	–
COAD	Colon adenocarcinoma	APOBEC1, DKC1, HENMT1	ADARB1
ESCA	Esophageal carcinoma	HENMT1, RBM15, ZC3H13	ADAD2, METTL17
GBM	Glioblastoma multiforme	APOBEC3G, HSD17B10	ADARB2, FBLL1, TARBP1
HNSC	Head and neck squamous cell carcinoma	HENMT1	–
KICH	Kidney chromophobe	ISCU	–
KIRC	Kidney renal clear cell carcinoma	APOBEC3G	–
KIRP	Kidney renal papillary cell carcinoma	METTL27	–
LAML	Acute myeloid leukemia	APOBEC3A, HENMT1	–
LGG	Brain lower-grade glioma	HSD17B10, TRMT5	ADARB1, HENMT1, TARBP1
LIHC	Liver hepatocellular carcinoma	LAGE3, METTL27, TARBP1	ADAT2, METTL17, NSUN6, TRMT11, ZC3H13
LUAD	Lung adenocarcinoma	LAGE3, METTL1, TFB2M	ADARB1, APOBEC3A, TRMT9B
LUSC	Lung squamous cell carcinoma	DKC1, FBL, LAGE3, METTL1, METTL8	ADARB1, TRMT9B
OV	Ovarian serous cystadenocarcinoma	–	YTHDC2
PAAD	Pancreatic adenocarcinoma	APOBEC1, APOBEC3G, HENMT1	QTRT1, WDR6
PCPG	Pheochromocytoma and paraganglioma	FBLL1	MPST
PRAD	Prostate adenocarcinoma	–	APOBEC3G, METTL17
READ	Rectum adenocarcinoma	APOBEC1, DKC1, HENMT1	ADARB1
SARC	Sarcoma	HENMT1	ADARB1, ISCU, SERGEF
SKCM	Skin cutaneous melanoma	APOBEC3G	–
STAD	Stomach adenocarcinoma	HENMT1	METTL17, TRMT2A, WDR6
TGCT	Testicular germ cell tumors	EMG1, FBL, HNRNPC, HSD17B10, LAGE3, METTL9	ADAD1, ADAD2
THCA	Thyroid carcinoma	LAGE3, TRMT112	ADAT2, METTL17, TRMT9B, TRMU
UCEC	Uterine corpus endometrial carcinoma	LAGE3	ADARB1
UCS	Uterine carcinosarcoma	FBL, HENMT1, HSD17B10, LAGE3, TRMT112	

Recent works have shown that specific RNA modifications are essential for the transmission of paternal diet-induced phenotypes intergenerationally [54]. Here, we identify two RMPs (TRDMT1 and METTL1) whose expression is significantly enriched in epididymis (Additional file 13: Figure S3), one of which (TRDMT1) was recently shown to be involved in the transmission of diet-induced paternal phenotypes across generations [54]. Whether METTL1 plays a role in intergenerational inheritance is yet to be deciphered; however, recent insights showing its role in miRNA maturation [61] suggest



**Fig. 5** Immunohistochemical analysis and prognostic value of RMP expression levels in different cancer types. **a, b** Immunohistochemical analysis and images of normal and tumor LAGE-3 stained LUSC (lung squamous cell carcinoma), LIHC (liver hepatocellular carcinoma), and PRAD (prostate adenocarcinoma) (a) and HENMT-1 stained HGSC (High-grade serous carcinoma), LUSC, and STAD (stomach adenocarcinoma) TMAs (b). Representative cores and subsets are shown for each tissue and antibody, where the brown color indicates a specific staining of the antibody and blue represents the hematoxylin counterstain. Mean TMA score is plotted for each core, with three cores from different individuals per condition quantified. Two-sided Wilcoxon tests did not yield significant differences in any comparison, *p* values of all tumor-normal comparisons for each cancer type and antibody are shown in Figure S13. **c** Heatmap of survival *p* values of 146 RMPs across 28 cancer types. Survival *p* values are calculated by comparing the prognosis of patients that express high (upper 50%) versus low (lower 50%) RMP levels. "N" column shows the number of patients included for the analysis of each cancer type. **d** Individual examples of survival plots where the expression levels of the RMP are predictive of cancer prognosis. *p* values have been calculated by comparing the survival between patients expressing high levels (yellow, top 50%) versus low expression levels (blue, bottom 50%)

that this enzyme might be playing a role in miRNA-acquired inheritance of information.

In the last few years, several studies have placed RNA modifications in the forefront of cancer research [36, 38, 68, 70, 83], mostly focused on the machinery responsible for writing and erasing m<sup>6</sup>A modifications. For many years, FTO was thought to be of special interest due to its association with obesity [84]. However, later studies proved this genome-wide association to be false [85] and that the single nucleotide variant present in the FTO intron was in fact associated with the activity of neighboring genes [85].

Nonetheless, FTO kept receiving special attention due to its perceived activity as an eraser of N6-methyladenosine (m<sup>6</sup>A) [27], the most frequent type of RNA modifications present in mRNAs. However, this is now thought to be incorrect, as later studies showed that FTO is in fact an eraser of N6,2'O-methyladenosine (m<sup>6</sup>Am), which is much less abundant in mRNAs [28, 86]. Similarly, FTO has been proposed to constitute a promising target for antitumor therapies [38, 80, 87]. While FTO has been shown to play an important role in leukemia [87], it is possible that additional RMPs such as HENMT1, which is drastically dysregulated in this cancer type, might constitute a better drug target to inhibit leukemogenesis (Fig. 4).

Here, we show that the expression of 40 RMPs is significantly altered in tumor samples, relative to their matched normal samples (Table 1 and Fig. 4). Moreover, we identify two enzymes, LAGE3 and HENMT1, as the top recurrently upregulated RMPs across cancer types. Surprisingly, these proteins have so far received little attention in cancer research studies. LAGE3 mutations are known to cause multiple human diseases, including nephrotic syndrome and microcephaly [88]; however, its role in tumorigenesis and cancer progression is yet to be determined. Here, we attempted to validate the upregulation of LAGE3 and HENMT1 across a battery of 12 cancer types using tissue microarrays (TMAs) (Fig. 5). While we were able to identify several cancer types where LAGE3 and HENMT1 were consistently upregulated, the variability among cancer grades across the tumor cores, together with the low number of cores per tumor type ( $n=3$ ) led to insufficient statistical power to identify significant expression changes. Future work will be needed to decipher the biological role of LAGE3 and HENMT1 in cancer, as well as its potential use as a target for diagnostic and prognostic purposes.

## Conclusions

Our analyses reveal an unanticipated heterogeneity in the expression patterns of RMPs across healthy mammalian tissues, with an over-representation of testis-specific RMPs, many of which are essential for sperm formation and maturation and, in some cases, are required for the transmission of epigenetic information across generations. In addition, we uncover a large proportion of dysregulated RMPs in multiple cancer types and show that several RMPs are dysregulated to a much larger extent than commonly studied m<sup>6</sup>A modification pathway, stressing the need to extend the epitranscriptomic drug targeting strategies to additional RNA modification enzymes. Now that novel transcriptome-wide tools to map additional RNA modifications have been recently made available, the community can repurpose antitumoral strategies to those RNA modification pathways that are most significantly dysregulated in each cancer type.

## Methods

### Compilation of human RNA modification-related proteins (RMPs)

An initial list of human methyltransferases, deaminases, and pseudouridylases was obtained by merging the lists available in the MODOMICS database (<http://modomics.genesilico.pl/>) and from a recently published review [13]. These lists were further initially completed with candidate genes by the addition of annotated proteins on Uniprot [42]. For each of these proteins, hidden Markov model (HMM) profiles of the corresponding PFAM catalytic domains were retrieved (Additional file 8: Table S8) by querying the PFAM database (<https://pfam.xfam.org/>). Each HMM profile was then used to query the human proteome using the `hmmsearch` function from HMMER software v.3.2.1 (<http://hmmer.org/>). Proteins above default threshold we kept as candidate RMW proteins (Additional file 9: Table S9). Related information for each of these proteins (modification type, target RNA, localization) was extracted from Uniprot, as well as from relevant literature [42]. Additional tRNA writer proteins were gathered from a recent study matching tRNA modifications to their writers [89]. Readers, erasers, and non-catalytic subunit proteins were obtained from annotated Uniprot genes as well as from published literature [90]. APOBEC3G and APOBEC3A were included in the analyses due to recent literature showing their deamination activity on RNA molecules in vivo in addition to acting on DNA [91, 92].

### Phylogenetic analysis

We first built a set of representative eukaryotic species, by choosing one species for each major phylogenetic clade for which complete proteomes were available. Our final list of representative species consisted of 25 complete proteomes from UniProt [42], which included 23 eukaryal species, as well as 2 outgroups (1 bacteria and 1 archaea) (Additional file 10: Table S10). For each proteome and RMW, we performed HMM-based searches, as described above. Candidate orthologs were manually curated to ensure that we did not miss any ortholog in our analysis, which resulted in a final table of RMW ortholog proteins (Additional file 11: Table S11). For each curated ortholog dataset, multiple sequence alignments were built using MAFFT with G-INS-1 method [93]. Alignment files were used to construct maximum likelihood phylogenetic trees using IQ-Tree with bootstrapping ( $n = 5000$ ) [94]. Consensus trees were visualized using FigTree v 1.4.4 [95] and used to identify the duplication events (Additional file 2: Table S2).

### Tissue specificity analysis

Human mRNA expression levels (TPM—transcripts per kilobase million) for each of the 146 human RMPs were downloaded from the Genotype Tissue Expression (GTEx) dataset [45], version v7, as well as from the Human Protein Atlas (HPA) [96]. Three GTEx tissues (whole blood, transformed lymphocytes and transformed fibroblasts) were discarded from downstream analyses, as these have been previously considered as outliers that can bias the analyses [45] or are not normal tissues of the human body. mRNA expression levels for adult mouse tissues (TPM) were obtained from a study that is part of the ENCODE project [97]. For each dataset (HPA, GTEx, ENCODE), we log transformed the TPM values after the addition of a pseudocount. To determine



which genes were tissue-specific, we compared the expression levels of RMP in a given tissue to the median expression levels of RMPs across all tissues. We then calculated residuals (using *rlm* function), which we refer to as “tissue specificity score” (TS), for each RMP to the regression line of each tissue. An RMP was considered tissue-specific if their TS was greater than 2.5 standard deviation (SD), as previously described [48], which, in a normal distribution of the standardized residuals, equals to the region outside of the 97.9 percentiles.

#### **RNA extraction from mice tissues and quantitative real-time PCR**

Brain, liver, lung, and testis tissues were collected from 20-week-old C57BL/6 J mice in triplicate. RNA was extracted from tissues using TRIzol™ Reagent (15596018, Thermo Fisher Scientific) and Chloroform (C2432, Vidra Foc) as per the manufacturer’s instructions, and precipitated with isopropanol (BP2618-500, Thermo Fisher Scientific) and Pellet Paint® Co-Precipitant (69049, Novagen). Samples were DNase treated with Turbo™ DNase (AM2238, Thermo Fisher Scientific) for 15 min at 37 °C and cleaned up using Agencourt RNAClean XP beads (A63987, Beckman Coulter) as per the manufacturer’s instructions. Quality of the extracted RNA was assessed using a Nanodrop™ Spectrophotometer 2000. cDNA was synthesized using Superscript II™ (18064014, Thermo Fisher Scientific) following the manufacturer’s instructions. Quantitative Real-Time PCR (qRT-PCR) was performed with Power SYBR™ Green PCR Mix (4367659, Thermo Fisher Scientific) using ViiA™ 7 Real-Time PCR System as per the manufacturer’s instructions. For each primer pair, three biological replicates with three technical replicate reactions were performed (total of 9 reactions per primer pair). *METTL5*, which is expressed stably among the four mouse tissues studied [98], was used for normalization purposes. Results were also analyzed using *GAPDH* for normalization purposes. qRT-PCR plots were built using GraphPad Prism 8. All oligonucleotides used for qRT-PCR can be found listed in Additional file 12: Table S12.

#### **RMP expression analysis across tissues in amniote species**

mRNA expression levels of 12 amniote species (human, chimpanzee, bonobo, gorilla, orangutan, rhesus macaque, mouse, gray-short tailed opossum, platypus, and chicken) were obtained from GSE30352 [99]. Normalized RPKM values of constitutive exons for both amniote and primate orthologs were used for downstream analyses. Heatmaps of the log transformed (with a pseudocount) and row (gene) *z*-scaled tissue-wide mRNA expression values were built using *complex heatmap* R package. PCA analysis was performed using *prcomp* function of R, and plots of scores (amniote and primate tissues) and loadings (orthologous genes) were plotted for the first two principal components using *ggplot* R package.

#### **Analysis of RMPs expression during spermatogenesis**

Processed spermatogenesis data was extracted from GSE112393 [55]. Input data was used to perform k-means clustering of RMPs based on their expression profiles in different sperm cell populations. The optimal number of clusters was calculated by plotting the within groups sum of squares by number of clusters extracted using k-means function in R, following criteria used by Scree’s test. Heatmaps were built using the

*complex heatmap* R package. Violin plots were built using the *ggplot* R package. To assess the consistency of our results across diverse datasets, we analyzed the RMP expression patterns from two additional mouse spermatogenesis studies [58, 59]. For the first dataset, we used the same gene cluster groups and plotted the corresponding heatmap and violin plots using the *ggplot* R package (Additional file 13: Figure S7). For the second dataset, we obtained the graphical representations for individual RMPs (Additional file 13: Figure S8) from the interactive website accompanying the paper [59].

### Immunohistochemistry

Testis and epididymis from 6- to 12-week-old C57BL/6J mice were fixed overnight at 4 °C with neutral buffered formalin (HT501128-4L, Sigma-Aldrich) and embedded in paraffin. Paraffin-embedded tissue sections (3 µm in thickness) were air dried and further dried at 60 °C overnight. Immunohistochemistry was performed using The Discovery XT Ventana Platform (Roche). Antigen retrieval was performed with Discovery CC1 buffer (950-500, Roche). Primary antibodies rabbit polyclonal anti-NSUN2 (20854-1-AP, Proteintech), rabbit polyclonal anti-NSUN7 (PA5-54257, Thermo Fisher Scientific), rabbit polyclonal anti-HENMT1 (PA5-55866, Thermo Fisher Scientific), and rabbit polyclonal anti-METTL14 (HPA038002, HPA038002) were diluted 1:1000, 1:100, 1:150, and 1:2000 respectively with EnVision FLEX Antibody Diluent (K800621, Dako, Agilent) and incubated for 60 min. Secondary antibody OmniMap anti-rabbit HRP (760-4311) was incubated for 20 min. Detection of the labeling was performed using the ChromoMAP DAB (760-159, Roche). Sections were counterstained with hematoxylin (760-2021, Roche) and mounted with Dako Toluene-Free Mounting Medium (CS705, Agilent) using a Dako CoverStainer (Agilent). Specificity of staining was confirmed with a rabbit IgG, polyclonal Isotype Control (ab27478, Abcam). Bright-field images were acquired with a NanoZoomer-2.0 HT C9600 digital scanner (Hamamatsu) equipped with a ×20 objective. All images were visualized with a gamma correction set at 1.8 in the image control panel of the NDP.view 2 U123888-01 software (Hamamatsu, Photonics, France). Mice samples were collected, prepared as paraffin blocks, sliced, and stained at the IRB Histopathology Facility. Negative controls for each antibody were also included, which showed no staining (Additional file 13: Figure S15). All IHC experiments were performed in biological triplicates.

### Immunofluorescence

Testis and epididymis from 12-week-old C57BL/6J mice were embedded in Tissue-Tek® O.C.T™ Compound (4583, Sakura) and 12-µm sagittal sections were mounted on SuperFrost™ microscope slides (12372098, Thermo Fisher Scientific). Tissue sections were defrosted, circled with a PAP pen (Z377821, Sigma-Aldrich), fixed in 4% PFA (28908, Thermo Fisher Scientific) for 10 min, and permeabilized in 0.5% Triton-X 100 for 30 min (T8787, Sigma-Aldrich). Subsequently, sections were blocked in 5% BSA (A7906, Sigma-Aldrich) for 45 min at room temperature and incubated in primary antibody in 5% BSA overnight at 4 °C. Primary antibodies were used at the following dilutions; 1:40 rabbit polyclonal anti-NSUN2 (20854-1-AP, Proteintech), 1:20 rabbit polyclonal anti-NSUN7 (PA5-55866, Thermo Fisher Scientific), 1:50 mouse monoclonal anti-DDX4 (ab27591, Abcam), 1:250 mouse monoclonal anti-Fibrillarin (38F3, Novus

Biologicals), and 2  $\mu\text{g}/\text{mL}$  IgG Isotype controls (G3A1 and 2791, Cell Signalling). Sections were incubated with 1:400 Alexa488-coupled anti-mouse (A-11001, Thermo Fisher Scientific) and Alexa555-coupled anti-rabbit (A-21429, Thermo Fisher Scientific) secondaries and counterstained with 1:10,000 Hoechst 33342 (H3570, Life Technologies) for 2 h at room temperature, then mounted with Fluoromount™ Aqueous Mounting Medium (F4680, Sigma-Aldrich). Prepared slides were imaged on a Leica TCS SPE using a 63X NA1.4 oil objective. Three  $1024 \times 1024$  representative regions of interest were imaged per testis ( $n = 3$ ) over a 3D stack (3–5  $\mu\text{m}$  depth with a z-step size of 1  $\mu\text{m}$ ), using a zoom factor of 2. All images were captured with a frame average of 4, with the exception of Hoechst which was imaged with a frame average of 2.

### Analysis of RMP expression in tumor-normal paired human datasets

TPM expression values were downloaded from the UCSC XENA Project, which contains the TCGA and GTEX RNA Seq data that is processed together to provide more reliable expression analysis with tumor and normal samples [73]. We discarded CHOL, THYM, and DLBC tumor-normal tissue pairs due to lack of proper control of normal tissue (low number of patients) in these cancer types. Data was transformed into  $\log_2(\text{TPM} + 1)$  for downstream analyses. For the  $\log_2(\text{FC})$  analyses, we calculated the difference between median  $\log_2$  expression levels between tumor and normal datasets, for each cancer type and RMP. For dysregulation analysis, we calculated the residuals (using *rlm* function in R) for all of the gene expression in a given tumor tissue and normal tissue pair, which has been previously termed as “dysregulation score” (DS) [48]. We set the threshold of significance DS at 2.5 standard deviations (SD) as previously described [48], which, in a normal distribution of standardized residuals, equals to the region outside of the 97.9 percentiles. We then extracted the dysregulation scores of the RMPs and used it for further downstream analyses. For heatmap representations, dysregulation scores were scaled and centered, and the final heatmap was built using *complex heatmap* R package. Scatter plots of median  $\log_2$  expression values for all genes in tumor-normal paired data were built using the *ggplot* R package, highlighting RMPs in black, significantly dysregulated RMPs in red (upregulated in tumor) and blue (downregulated in tumor), and non-RMP proteins were depicted in gray.

### Survival analyses

Survival phenotypes were downloaded from the XENA Platform, using the “TCGA TARGET GTEX” cohort [73]. In order to analyze the survival data, we first determined patients that have “high” (upper 50% relative to average expression) and “low” (lower 50% relative to average expression) expression of a specific gene, and matched these patients with their overall survival information. We then used the *survminer* R package to plot survival curves for each gene and every cancer type, as well as to extract the survival  $p$  values.  $p$  values were transformed by inversion and subsequent log-transformation with a pseudocount [ $\log(1/p + 1)$ ]. Heatmap of the survival  $p$  values was built using *complex heatmap* R package. Transformed survival  $p$  values were visualized using *ggplot*.

### Tumor microarray immunohistochemistry and analysis

Multiorgan tumor with adjacent normal tissue microarray slides with accompanying pathology grade, TNM (tumor, node, and metastasis) classification, and clinical stage information was purchased from US Biomax Inc. (BCN721a). Each slide contains three malignant and three normal cores from 12 types of human organs (esophagus, stomach, colon, rectum, liver, lung, kidney, breast, cervix, ovary, prostate, and pancreas), each core taken from different individuals. TMAs were stained at IRB Histopathology Facility. Primary antibodies rabbit polyclonal anti-HENMT1 (PA5-55866, Thermo Fisher Scientific) and rabbit polyclonal anti-LAGE3 (HPA036123, Sigma-Aldrich) were diluted to 1:50. Secondary antibody OmniMap anti-rabbit HRP (760-4311) was incubated for 20 min. Detection of the labeling was performed using the ChromoMAP DAB (760-159, Roche). For scoring of tissue microarrays, each core was given a score from 0 to 4 based on the proportion of positively stained cells: 0 represents < 2% of cells staining positive, 1 represents 2–25%, 2 represents 26–50%, 3 represents 51–75%, and 4 represents 76–100% staining of cells [100]. Two blinded independent people scored the stainings, following the guidelines described above. Both scorers were blind to both the antibody and tissue type. The scores from each scorer were averaged to obtain the final score per core. A two-sided Wilcoxon test was used to assess significance.

### Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s13059-020-02009-z>.

**Additional file 1: Table S1.** List of human RNA modification-related proteins (RMPs) used in this study.

**Additional file 2: Table S2.** Duplication events of Catalytic RNA Writer Proteins (RMWs) used in this study.

**Additional file 3: Table S3.** Tissue-specific RNA modification-related proteins (RMPs) in human and mouse.

**Additional file 4: Table S4.** Duplication events of Catalytic RNA Writer Proteins (RMWs) with tissue and target specificity information.

**Additional file 5: Table S5.** Number of samples analyzed for each cancer type, both in Normal and Tumor tissues.

**Additional file 6: Table S6.** Dysregulation scores of significantly dysregulated RMPs.

**Additional file 7: Table S7.** Survival p-values of RMPs whose expression is significantly associated with survival.

**Additional file 8: Table S8.** PFAM domains used in phylogenetic analysis.

**Additional file 9: Table S9.** Raw output of Hmmssearch including RMWs and non-RMPs.

**Additional file 10: Table S10.** List of representative species used for phylogenetic analysis and manual curation of ortholog genes.

**Additional file 11: Table S11.** Uniprot list of the ortholog main catalytic RNA writer proteins in the representative species.

**Additional file 12: Table S12.** Primers used for qPCR.

**Additional file 13: Figure S1.** Expression analysis plots (Heatmap and PCA) of RMPs in Human and Mouse tissues. **Figure S2.** Quantitative real-time PCR of 8 RMPs expressed in four mouse tissues. **Figure S3.** Proteomics analysis of RMPs in human tissues. **Figure S4.** Expression of RMPs in Amniote and Primate species. **Figure S5.** Analysis of target specificity of tissue-specific and non-tissue-specific genes. **Figure S6.** Expression analysis of RMPs in mouse spermatogenesis and Immunohistochemical staining of HENMT1 in mouse testis and epididymis. **Figure S7.** Comparison of RMP expression changes during spermatogenesis using published single-cell RNA sequencing datasets (Green et al., 2018 and Xia et al., 2020). **Figure S8.** Comparison of RMP expression patterns during spermatogenesis, using the data published by Green et al., 2018 and Jung & Wells et al., 2019. **Figure S9.** Comparative analysis of mRNA expression levels of HENMT1, NSUN2, NSUN7 and METTL14 during spermatogenesis, extracted from 3 distinct single-cell RNAseq publicly available datasets. **Figure S10.** Immunofluorescence of NSUN2 and NSUN7 RMPs in mouse testis. **Figure S11.** Heatmap of the RMP expression changes (log<sub>2</sub>FC) between tumor and normal samples, across 28 cancer types. **Figure S12.** Scatterplots showing expression levels of RMPs in matched tumor-normal samples for all 28 cancer types analyzed. **Figure S13.** Tumor stage-specific RNA expression levels of LAGE3 and HENMT1. **Figure S14.** Immunohistochemical staining of Tissue microarray (TMA) with LAGE3 and HENMT1 antibodies. **Figure S15.** Immunohistochemical staining of mouse testis and epididymis using isotype control rabbit IgG antibody (negative control).

**Additional file 14.** Review history.

### Acknowledgements

We thank all the members of the Novoa lab for their valuable insights and discussion. The results shown here are in whole or part based upon data generated by the TCGA Research Network. We thank Anais Zaratzian for the initial setup of the immunohistochemistry experiments in the Histopathology Facility at the Garvan Institute. We also thank IRB Histopathology Facility and CRG Histology Unit for the immunohistochemistry experiments and tissue sectioning.

### Peer review information

Barbara Cheifet was the primary editor of this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

### Review history

The review history is available as Additional file 14.

### Funding

OB is supported by a UNSW International PhD fellowship. MCL is supported by an FPI Severo Ochoa PhD fellowship from the Spanish Ministry of Economy, Industry and Competitiveness (MEIC). EMN was supported by a Discovery Early Career Researcher Award (DE170100506) from the Australian Research Council and is currently supported by CRG Severo Ochoa Funding. This work was supported by the Australian Research Council (DP180103571 to EMN), by the Spanish Ministry of Economy, Industry and Competitiveness (PGC2018-098152-A-100 to EMN) and NHMRC funds (Project Grant APP1070631 to JSM). We acknowledge the support of the Spanish Ministry of Economy, Industry and Competitiveness (MEIC) to the EMBL partnership, Centro de Excelencia Severo Ochoa, and CERCA Programme / Generalitat de Catalunya.

### Availability of data and materials

All scripts used in this work have been made publicly available and can be found at <https://github.com/novoalab/RNAModMachinery> [101]. All datasets used to build the figures, as well as intermediate analysis files (alignment files, maximum likelihood trees, scatter plots of tissue specificity, barplots of amniote and primate ortholog expressions, scatter plots of tumor vs normal tissues, boxplots of individual expression of RMPs in tumor-normal paired tissues and survival plots), are publicly available at [https://public-docs.crg.es/enovoa/public/website/Begik\\_RMP2020.html](https://public-docs.crg.es/enovoa/public/website/Begik_RMP2020.html). Raw immunofluorescence images and IHC scans have been deposited in Figshare [102, 103].

Third-party mRNA expression data used throughout this work were obtained from the following resources: (i) mRNA expression datasets across human tissues were obtained from GTEx (<https://gtexportal.org/home/index.html>) [45] and HPA (<https://www.proteinatlas.org/>) [96]; (ii) mRNA expression datasets for mouse tissues were obtained from ENCODE (<https://www.encodeproject.org/>) [97]; (iii) mRNA expression levels across tissues from 12 amniote species were obtained from GSE30352 [99]; (iv) single-cell RNASeq levels during mouse spermatogenesis was obtained from GSE112393 [55], GSE125372 [58, 59], and GSE113293 [59]; (v) mRNA expression data from tumor-normal human samples were downloaded from the UCSC XENA Project (<https://xenabrowser.net/>) [73]; (vi) survival phenotypes were downloaded from the XENA Platform (<https://xenabrowser.net/>), using the "TCGA TARGET GTEx" cohort [73].

### Authors' contributions

OB performed all bioinformatic analyses. OB and MCL performed mouse tissue RNA extraction and qRT-PCR experiments. MCL performed and analyzed IF experiments. MCL and OB scored and analyzed the TMAs. MCL contributed to IHC experiments as well as to the corresponding figures. HL and JMR contributed to code and suggestions for the data analysis. OB made all figures, with the contribution from MCL and EMN. OB, MCL, and EMN interpreted the results. EMN conceived the project. EMN supervised the work, with the contribution of JSM. OB, EMN, and MCL wrote the manuscript, with input from all authors. The authors read and approved the final manuscript.

### Ethics approval and consent to participate

Experiments for tissue collection from C57BL/6 mice used for IHC were approved by the Generalitat de Catalunya and were conducted under CEEA Ethics Approval protocols 9162-P2 and 9350.

### Competing interests

The authors declare that they have no competing interests.

### Author details

<sup>1</sup>Center for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, 08003 Barcelona, Spain.

<sup>2</sup>Garvan Institute of Medical Research, Darlinghurst, NSW 2010, Australia. <sup>3</sup>UNSW, Sydney, Sydney, NSW 2052, Australia.

<sup>4</sup>Universitat Pompeu Fabra (UPF), Barcelona, Spain.

Received: 29 October 2019 Accepted: 3 April 2020

Published online: 07 May 2020

### References

1. Saletore Y, Meyer K, Korlach J, Vilfan ID, Jaffrey S, Mason CE. The birth of the epitranscriptome: deciphering the function of RNA modifications. *Genome Biol.* 2012;13:175.
2. Zhou J, Wan J, Shu XE, Mao Y, Liu X-M, Yuan X, et al. N6-methyladenosine guides mRNA alternative translation during integrated stress response. *Mol Cell.* 2018;69:636–47.e7.
3. Safra M, Sas-Chen A, Nir R, Winkler R, Nachshon A, Bar-Yaacov D, et al. The m1A landscape on cytosolic and mitochondrial mRNA at single-base resolution. *Nature.* 2017;551:251–5.

4. Warda AS, Kretschmer J, Hackert P, Lenz C, Urlaub H, Höbartner C, et al. Human METTL16 is a N6-methyladenosine (m6A) methyltransferase that targets pre-mRNAs and various non-coding RNAs. *EMBO Rep.* 2017;18:2004–14.
5. Lim SL, Qu ZP, Kortschak RD, Lawrence DM, Geoghegan J, Hempfling A-L, et al. HENMT1 and piRNA stability are required for adult male germ cell transposon repression and to define the spermatogenic program in the mouse. *PLoS Genet.* 2015;11:e1005620.
6. Shoshan E, Mobley AK, Braeuer RR, Kamiya T, Huang L, Vasquez ME, et al. Reduced adenosine-to-inosine miR-455-5p editing promotes melanoma growth and metastasis. *Nat Cell Biol.* 2015;17:311–21.
7. Batista PJ, Molinie B, Wang J, Qu K, Zhang J, Li L, et al. m(6)A RNA modification controls cell fate transition in mammalian embryonic stem cells. *Cell Stem Cell.* 2014;15:707–19.
8. Kan L, Grozhik AV, Vedanayagam J, Patil DP, Pang N, Lim K-S, et al. The m6A pathway facilitates sex determination in *Drosophila*. *Nat Commun.* 2017;8:15737.
9. Lence T, Akhtar J, Bayer M, Schmid K, Spindler L, Ho CH, et al. m6A modulates neuronal functions and sex determination in *Drosophila*. *Nature.* 2016;540:242–7.
10. Zhao BS, Wang X, Beadell AV, Lu Z, Shi H, Kuuspalu A, et al. m(6)A-dependent maternal mRNA clearance facilitates zebrafish maternal-to-zygotic transition. *Nature.* 2017;542:475–8.
11. Fustin JM, Doi M, Yamaguchi Y, Hida H, Nishimura S, Yoshida M, et al. RNA-methylation-dependent RNA processing controls the speed of the circadian clock. *Cell.* 2013;155:793–806.
12. Vandivier LE, Gregory BD. New insights into the plant epitranscriptome. *J Exp Bot.* 2018;69:4659–65.
13. Jonkhout N, Tran J, Smith MA, Schonrock N, Mattick JS, Novoa EM. The RNA modification landscape in human disease. *RNA.* 2017;23:1754–69.
14. Torres AG, Batlle E, Ribas de Pouplana L. Role of tRNA modifications in human diseases. *Trends Mol Med.* 2014;20:306–14.
15. Bednarova A, Hanna M, Durham I, VanCleave T, England A, Chaudhuri A, et al. Lost in translation: defects in transfer RNA modifications and neurological disorders. *Front Mol Neurosci.* 2017;10:135.
16. Sarin LP, Leidel SA. Modify or die?—RNA modification defects in metazoans. *RNA Biol.* 2014;11:1555–67.
17. Pereira M, Francisco S, Varanda AS, Santos M, Santos MAS, Soares AR. Impact of tRNA modifications and tRNA-modifying enzymes on proteostasis and human disease. *Int J Mol Sci.* 2018;19 <https://doi.org/10.3390/ijms19123738>.
18. Alexandrov A, Chernyakov I, Gu W, Hiley SL, Hughes TR, Grayhack EJ, et al. Rapid tRNA decay can result from lack of nonessential modifications. *Mol Cell.* 2006;21:87–96.
19. Schaefer M, Pollex T, Hanna K, Tuorto F, Meusburger M, Helm M, et al. RNA methylation by Dnmt2 protects transfer RNAs against stress-induced cleavage. *Genes Dev.* 2010;24:1590–5.
20. Ke S, Pandya-Jones A, Saito Y, Fak JJ, Vågbo CB, Geula S, et al. m6A mRNA modifications are deposited in nascent pre-mRNA and are not required for splicing but do specify cytoplasmic turnover. *Genes Dev.* 2017;31:990–1006.
21. Novoa EM, Ribas de Pouplana L. Speeding with control: codon usage, tRNAs, and ribosomes. *Trends Genet.* 2012;28:574–81.
22. Wang X, Zhao BS, Roundtree IA, Lu Z, Han D, Ma H, et al. N(6)-methyladenosine Modulates Messenger RNA Translation Efficiency. *Cell.* 2015;161:1388–99.
23. Kaneko T, Suzuki T, Kapushoc ST, Rubio MA, Ghazvini J, Watanabe K, et al. Wobble modification differences and subcellular localization of tRNAs in *Leishmania tarentolae*: implication for tRNA sorting mechanism. *EMBO J.* 2003;22:657–67.
24. Dal Magro C, Keller P, Kotter A, Werner S, Duarte V, Marchand V, et al. A vastly increased chemical variety of RNA modifications containing a thioacetal structure. *Angew Chem Int Ed Engl.* 2018;57:7893–7.
25. Bokar JA, Shambaugh ME, Polayes D, Matera AG, Rottman FM. Purification and cDNA cloning of the AdoMet-binding subunit of the human mRNA (N6-adenosine)-methyltransferase. *RNA.* 1997;3:1233–47.
26. Shi H, Wei J, He C. Where, when, and how: context-dependent functions of RNA methylation writers, readers, and erasers. *Mol Cell.* 2019;74:640–50.
27. Jia G, Fu Y, Zhao X, Dai Q, Zheng G, Yang Y, et al. N6-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. *Nat Chem Biol.* 2011;7:885–7.
28. Mauer J, Luo X, Blanjoie A, Jiao X, Grozhik AV, Patil DP, et al. Reversible methylation of m6Am in the 5' cap controls mRNA stability. *Nature.* 2016; Available from: <https://www.ncbi.nlm.nih.gov/pubmed/28002401>. Accessed 10 May 2019.
29. Louloui A, Ntini E, Conrad T, Orom UAV. Transient N-6-methyladenosine transcriptome sequencing reveals a regulatory role of m6A in splicing efficiency. *Cell Rep.* 2018;23:3429–37.
30. Kasowitz SD, Ma J, Anderson SJ, Leu NA, Xu Y, Gregory BD, et al. Nuclear m6A reader YTHDC1 regulates alternative polyadenylation and splicing during mouse oocyte development. *PLoS Genet.* 2018;14:e1007412.
31. Tang C, Klukovich R, Peng H, Wang Z, Yu T, Zhang Y, et al. ALKBH5-dependent m6A demethylation controls splicing and stability of long 3'-UTR mRNAs in male germ cells. *Proc Natl Acad Sci U S A.* 2018;115:E325–33.
32. Li A, Chen Y-S, Ping X-L, Yang X, Xiao W, Yang Y, et al. Cytoplasmic m6A reader YTHDF3 promotes mRNA translation. *Cell Res.* 2017;27:444–7.
33. Yu J, Chen M, Huang H, Zhu J, Song H, Zhu J, et al. Dynamic m6A modification regulates local translation of mRNA in axons. *Nucleic Acids Res.* 2018;46:1412–23.
34. Lin Z, Hsu PJ, Xing X, Fang J, Lu Z, Zou Q, et al. Mettl3-/Mettl14-mediated mRNA N6-methyladenosine modulates murine spermatogenesis. *Cell Res.* 2017;27:1216–30.
35. Lin Z, Tong M-H. m6A mRNA modification regulates mammalian spermatogenesis. *Biochim Biophys Acta Gene Regul Mech.* 2018; <https://doi.org/10.1016/j.bbagr.2018.10.016>.
36. Vu LP, Pickering BF, Cheng Y, Zaccara S, Nguyen D, Minuesa G, et al. The N6-methyladenosine (m6A)-forming enzyme METTL3 controls myeloid differentiation of normal hematopoietic and leukemia cells. *Nat Med.* 2017;23:1369–76.
37. Lian H, Wang Q-H, Zhu C-B, Ma J, Jin W-L. Deciphering the epitranscriptome in cancer. *Trends Cancer Res.* 2018;4:207–21.
38. Lan Q, Liu PY, Haase J, Bell JL, Hüttelmaier S, Liu T. The critical role of RNA m6A methylation in cancer. *Cancer Res.* 2019;79:1285–92.
39. Lin X, Chai G, Wu Y, Li J, Chen F, Liu J, et al. RNA m6A methylation regulates the epithelial mesenchymal transition of cancer cells and translation of snail. *Nat Commun.* 2019;10:2065.

40. Jaffrey SR, Kharas MG. Emerging links between m6A and misregulated mRNA methylation in cancer. *Genome Med.* 2017;9:2.
41. Boccaletto P, Machnicka MA, Purta E, Piatkowski P, Baginski B, Wirecki TK, et al. MODOMICS: a database of RNA modification pathways. 2017 update. *Nucleic Acids Res.* 2018;46:D303–7.
42. UniProt Consortium. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* 2019;47:D506–15.
43. Xu L, Liu X, Sheng N, Oo KS, Liang J, Chionh YH, et al. Three distinct 3-methylcytidine (m(3)C) methyltransferases modify tRNA and mRNA in mice and humans. *J Biol Chem.* 2017;292:14695–703.
44. Vilardo E, Nachbagauer C, Buzet A, Taschner A, Holzmann J, Rossmannith W. A subcomplex of human mitochondrial RNase P is a bifunctional methyltransferase—extensive moonlighting in mitochondrial tRNA biogenesis. *Nucleic Acids Res.* 2012;40:11583–93.
45. GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science.* 2015;348:648–60.
46. Thul PJ, Lindskog C. The human protein atlas: a spatial map of the human proteome. *Protein Sci.* 2018;233–44. <https://doi.org/10.1002/pro.3307>.
47. Li B, Qing T, Zhu J, Wen Z, Yu Y, Fukumura R, et al. A comprehensive mouse transcriptomic BodyMap across 17 tissues by RNA-seq. *Sci Rep.* 2017;7:4200.
48. Guimaraes JC, Zavolan M. Patterns of ribosomal protein expression specify normal and malignant human cells. *Genome Biol.* 2016;17:236.
49. Kim M-S, Pinto SM, Getnet D, Nirujogi RS, Manda SS, Chaerkady R, et al. A draft map of the human proteome. *Nature.* 2014;509:575–81.
50. Chen Y, Zheng Y, Gao Y, Lin Z, Yang S, Wang T, et al. Single-cell RNA-seq uncovers dynamic processes and critical regulators in mouse spermatogenesis. *Cell Res.* 2018;28:879–96.
51. Bettgowda A, Wilkinson MF. Transcription and post-transcriptional regulation of spermatogenesis. *Philos Trans R Soc Lond B Biol Sci.* 2010;365:1637–51.
52. Robles V, Herráez P, Labbé C, Cabrera E, Pšenička M, Valcarce DG, et al. Molecular basis of spermatogenesis and sperm quality. *Gen Comp Endocrinol.* 2017;245:5–9.
53. Jiang J, White-Cooper H. Transcriptional activation in *Drosophila* spermatogenesis involves the mutually dependent function of aly and a novel meiotic arrest gene cookie monster. *Development.* 2003;130:563–73.
54. Zhang Y, Zhang X, Shi J, Tuorto F, Li X, Liu Y, et al. Dnmt2 mediates intergenerational transmission of paternally acquired metabolic disorders through sperm small non-coding RNAs. *Nat Cell Biol.* 2018;20:535–40.
55. Green CD, Ma Q, Manske GL, Shami AN, Zheng X, Marini S, et al. A comprehensive roadmap of murine spermatogenesis defined by single-cell RNA-Seq. *Dev Cell.* 2018;46:651–67.e10.
56. Connolly CM, Dearth AT, Braun RE. Disruption of murine Tenr results in teratospermia and male infertility. *Dev Biol.* 2005; 278:13–21.
57. Harris T, Marquez B, Suarez S, Schimenti J. Sperm motility defects and infertility in male mice with a mutation in Nsun7, a member of the Sun domain-containing family of putative RNA methyltransferases. *Biol Reprod.* 2007;77:376–82.
58. Xia B, Yan Y, Baron F, Wagner F, Barkley D, Chiodin M, et al. Widespread transcriptional scanning in the testis modulates gene evolution rates. *Cell.* 2020;180:248–62.e21.
59. Jung M, Wells D, Rusch J, Ahmad S, Marchini J, Myers SR, et al. Unified single-cell analysis of testis gene regulation and pathology in five mouse strains. *Elife.* 2019;8 <https://doi.org/10.7554/eLife.43966>.
60. Tuorto F, Liebers R, Musch T, Schaefer M, Hofmann S, Kellner S, et al. RNA cytosine methylation by Dnmt2 and NSun2 promotes tRNA stability and protein synthesis. *Nat Struct Mol Biol.* 2012;19:900–5.
61. Pandolfini L, Barbieri I, Bannister AJ, Hendrick A, Andrews B, Webster N, et al. METTL1 Promotes let-7 MicroRNA Processing via m7G Methylation. *Mol Cell.* 2019;74:1278–90.e9.
62. Vogel C, Marcotte EM. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat Rev Genet.* 2012;13:227–32.
63. Khosronezhad N, Hosseinzadeh Colagar A, Mortazavi SM. The Nsun7 (A11337)-deletion mutation, causes reduction of its protein rate and associated with sperm motility defect in infertile men. *J Assist Reprod Genet.* 2015;32:807–15.
64. Hussain S, Tuorto F, Menon S, Blanco S, Cox C, Flores JV, et al. The mouse cytosine-5 RNA methyltransferase NSun2 is a component of the chromatoid body and required for testis differentiation. *Mol Cell Biol.* 2013;33:1561–70.
65. Mathieu C, Guérin JF, Cognat M, Lejeune H, Pinatel MC, Lornage J. Motility and fertilizing capacity of epididymal human spermatozoa in normal and pathological cases. *Fertil Steril.* 1992;57:871–6.
66. Wolfson B, Gambone J, Rajfer J. Identification of motile sperm in caput epididymis. Intraoperative observations and clinical correlations. *Urology.* 1992;40:335–8.
67. Kotaja N, Bhattacharyya SN, Jaskiewicz L, Kimmins S, Parvinen M, Filipowicz W, et al. The chromatoid body of male germ cells: similarity with processing bodies and presence of Dicer and microRNA pathway components. *Proc Natl Acad Sci U S A.* 2006;103:2647–52.
68. Delaunay S, Frye M. RNA modifications regulating cell fate in cancer. *Nat Cell Biol.* 2019;21:552–9.
69. Cui Q, Shi H, Ye P, Li L, Qu Q, Sun G, et al. m6A RNA methylation regulates the self-renewal and tumorigenesis of glioblastoma stem cells. *Cell Rep.* 2017;18:2622–34.
70. Paris J, Morgan M, Campos J, Spencer GJ, Shmakova A, Ivanova I, et al. Targeting the RNA m6A reader YTHDF2 selectively compromises cancer stem cells in acute myeloid leukemia. *Cell Stem Cell.* 2019; <https://doi.org/10.1016/j.stem.2019.03.021>.
71. Pinello N, Sun S, Wong JJ-L. Aberrant expression of enzymes regulating m6A mRNA methylation: implication in cancer. *Cancer Biol Med.* 2018;15:323–34.
72. Okamoto M, Fujiwara M, Hori M, Okada K, Yazama F, Konishi H, et al. tRNA modifying enzymes, NSUN2 and METTL1, determine sensitivity to 5-fluorouracil in HeLa cells. *PLoS Genet.* 2014;10:e1004639.
73. Goldman M, Craft B, Hastie M, Repecka K, Kamath A, McDade F, et al. The UCSC Xena platform for public and private cancer genomics data visualization and interpretation. *bioRxiv.* 2019:326470.
74. Vagin VV, Sigova A, Li C, Seitz H, Gvozdev V, Zamore PD. A distinct small RNA pathway silences selfish genetic elements in the germline. *Science.* 2006;313:320–4.

75. Kirino Y, Mourelatos Z. 2'-O-methyl modification in mouse piRNAs and its methylase. *Nucleic Acids Symp Ser.* 2007;51:417–8.
76. Thiaville PC, El Yacoubi B, Köhrer C. Essentiality of threonylcarbamoyladenine (t6A), a universal tRNA modification, in bacteria. *Molecular.* 2015; Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1111/mmi.13209>. Accessed 1 Nov 2019.
77. Thiaville PC, Legendre R, Rojas-Benítez D, Baudin-Baillieu A, Hatin I, Chalancon G, et al. Global translational impacts of the loss of the tRNA modification t6A in yeast. *Microb Cell Fact.* 2016;3:29–45.
78. Janin M, Ortiz-Barahona V, de Moura MC, Martínez-Cardús A, Llinàs-Arias P, Soler M, et al. Epigenetic loss of RNA-methyltransferase NSUN5 in glioma targets ribosomes to drive a stress adaptive translational program. *Acta Neuropathol.* 2019; <https://doi.org/10.1007/s00401-019-02062-4>.
79. Öunap K, Käsper L, Kurg A, Kurg R. The human WBSR22 protein is involved in the biogenesis of the 40S ribosomal subunits in mammalian cells. *PLoS One.* 2013;8:e75686.
80. Huang Y, Su R, Sheng Y, Dong L, Dong Z, Xu H, et al. Small-molecule targeting of oncogenic FTO demethylase in acute myeloid leukemia. *Cancer Cell.* 2019;35:677–91.e10.
81. Guschanski K, Warnefors M, Kaessmann H. The evolution of duplicate gene expression in mammalian organs. *Genome Res.* 2017;27:1461–74.
82. Roundtree IA, Evans ME, Pan T, He C. Dynamic RNA modifications in gene expression regulation. *Cell.* 2017;169:1187–200.
83. Kato T, Daigo Y, Hayama S, Ishikawa N, Yamabuki T, Ito T, et al. A novel human tRNA-dihydrouridine synthase involved in pulmonary carcinogenesis. *Cancer Res.* 2005;65:5638–46.
84. Fawcett KA, Barroso I. The genetics of obesity: FTO leads the way. *Trends Genet.* 2010;26:266–74.
85. Claussnitzer M, Dankel SN, Kim K-H, Quon G, Meuleman W, Haugen C, et al. FTO obesity variant circuitry and adipocyte browning in humans. *N Engl J Med.* 2015;373:895–907.
86. Mauer J, Sindelar M, Despic V, Guez T, Hawley BR, Vasseur J-J, et al. FTO controls reversible m6Am RNA methylation during snRNA biogenesis. *Nat Chem Biol.* 2019;15:340–7.
87. Li Z, Weng H, Su R, Weng X, Zuo Z, Li C, et al. FTO plays an oncogenic role in acute myeloid leukemia as a N6-methyladenosine RNA demethylase. *cancer cell.* 2017;127–41. <https://doi.org/10.1016/j.ccell.2016.11.017>.
88. Braun DA, Rao J, Mollet G, Schapiro D, Daugeron M-C, Tan W, et al. Mutations in KEOPS-complex genes cause nephrotic syndrome with primary microcephaly. *Nat Genet.* 2017;49:1529–38.
89. de Crécy-Lagard V, Boccaletto P, Mangleburg CG, Sharma P, Lowe TM, Leidel SA, et al. Matching tRNA modifications in humans to their known and predicted enzymes. *Nucleic Acids Res.* 2019; <https://doi.org/10.1093/nar/gkz011>.
90. Lobo J, Costa AL, Cantante M, Guimarães R, Lopes P, Antunes L, et al. m6A RNA modification and its writer/reader VIRMA/YTHDF3 in testicular germ cell tumors: a role in seminoma phenotype maintenance. *J Transl Med.* 2019;17:79.
91. Sharma S, Patnaik SK, Taggart RT, Baysal BE. The double-domain cytidine deaminase APOBEC3G is a cellular site-specific RNA editing enzyme. *Sci Rep.* 2016;6:39100.
92. Sharma S, Patnaik SK, Taggart RT, Kannisto ED, Enriquez SM, Gollnick P, et al. APOBEC3A cytidine deaminase induces RNA editing in monocytes and macrophages. *Nat Commun.* 2015;6:6881.
93. Katoh K, Rozewicki J, Yamada KD. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief Bioinform.* 2017; <https://doi.org/10.1093/bib/bbx108>.
94. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol.* 2015;32:268–74.
95. Rambaut A. FigTree v1; 2012. p. 4.
96. Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Proteomics. Tissue-based map of the human proteome. *Science.* 2015;347:1260419.
97. Pervouchine DD, Djebali S, Breschi A, Davis CA, Barja PP, Dobin A, et al. Enhanced transcriptome maps from multiple mouse tissues reveal evolutionary constraint in gene expression. *Nat Commun.* 2015;6:5903.
98. Yue F, Cheng Y, Breschi A, Vierstra J, Wu W, Ryba T, et al. A comparative encyclopedia of DNA elements in the mouse genome. *Nature.* 2014;515:355–64.
99. Brawand D, Soumillon M, Necsulea A, Julien P, Csárdi G, Harrigan P, et al. The evolution of gene expression levels in mammalian organs. *Nature.* 2011;478:343–8.
100. Griffin MC, Robinson RA, Trask DK. Validation of tissue microarrays using p53 immunohistochemical studies of squamous cell carcinoma of the larynx. *Mod Pathol.* 2003;16:1181–8.
101. Begik O, Lucas MC, Liu H, Ramirez JM, Mattick JS, Nova EM. RNAModMachinery. Github. 2020; Available from: <https://github.com/novoalab/RNAModMachinery>. [cited 2020 Mar 27].
102. Begik O, Lucas MC, Liu H, Ramirez JM, Mattick JS, Nova EM. Immunofluorescence images. Figshare. 2020; <https://doi.org/10.6084/m9.figshare.12036762.v1>.
103. Begik O, Lucas MC, Liu H, Ramirez JM, Mattick JS, Nova EM. Immunohistochemistry images. FigShare. 2020; <https://doi.org/10.6084/m9.figshare.12036447.v1>.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.