

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

Oguzhan Begik^{1,2,3}, Morghan C. Lucas^{1,4}, Huanle Liu¹, Jose Miguel Ramirez^{1,4},

John S. Mattick^{2,3} and Eva Maria Novoa^{1,2,3,4*}

¹ *Center for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, Barcelona 08003, Spain.*

² *Garvan Institute of Medical Research, Darlinghurst, 2010, NSW, Australia.*

³ *UNSW, Sydney, Sydney 2052, NSW, Australia.*

⁴ *Universitat Pompeu Fabra (UPF), Barcelona, Spain.*

* Correspondence to: Eva Maria Novoa (eva.novoa@crg.eu)

ABSTRACT

Background: RNA modifications play central roles in cellular fate and differentiation. These features have placed the epitranscriptome in the forefront of developmental biology and cancer research. 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 have performed 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. Moreover, through the analysis of paired tumor-normal human samples 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 up-regulated in most cancer types, once the sample is properly scaled and normalized to the full dataset, 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 behaviours 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 post-transcriptional removal (by deamination, often called 'RNA editing') or by 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], amongst 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 (m⁶A), the most prevalent internal RNA modification found in human mRNAs. M⁶A is placed by a multicomponent transferase complex, in which methyltransferase 3 (METTL3) acts as the catalytic subunit [25,26]. Moreover, m⁶A modifications can be reversed by the activity of m⁶A 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 m⁶A marks [28]. Mechanistically, m⁶A 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 m⁶A 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 (**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 (**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 annotated human RMWs in other resources [41]. To determine the evolutionary history and identify duplication events occurred in each family, ortholog proteins from representative species were retrieved (see *Methods*), and phylogenetic trees were built to identify the number of duplications occurred within each family. Overall, our analysis identified 46 duplication events (**Figure 1A**), which have mainly occurred in the base of Eukaryota, Metazoa and Vertebrata (**Figure 1B**, see also **Table S2**).

We find that duplications are often accompanied by changes in substrate specificity (**Figures 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] (**Figure 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 (**Figure 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 - and possibly, different functions - 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] (**Figure 2A**, see also **Figures S1A** for gene-labelled 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 (**Figure 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 did not display none or few tissue-enriched RMPs (**Figure 2B**, see also **Figure S1B**). Moreover, we found that RMP tissue expression patterns were largely conserved in both mouse and human (**Figure 2C**, see also **Table S3**),

We then extended our analysis to additional amniote species, finding that testis was also the main outlier in terms of RMP expression patterns in all species analyzed, supporting the notion that testis-specific RMP functionalities are conserved (**Figure 2D**, see also **Figure S2**). 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%) (**Figure 2E**, see also **Table S4**), 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 (**Figure 3A**), is a highly-specialized differentiation process in which transcriptional, post-transcriptional and translational regulation are highly orchestrated [49–52]. 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 m6A depletion in mice dysregulates translation of transcripts that are required for spermatogonial proliferation and differentiation [34]. Moreover, m⁵C modifications have been shown to be essential for transmission of diet-induced epigenetic information across generations in the epididymis [53]. 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 [54] (**Figure 3B**, see also **Figure S3** for gene-labelled heatmap). We first classified RMPs based on their gene expression patterns (see *Methods*), identifying four main expression patterns: (i) high expression only during meiotic stages (spermatocytes and spermatids); (ii) high expression only during mitotic stages (spermatogonia); (iii) high expression in both mitotic and meiotic stages (spermatogonia, spermatocytes, spermatids), although decreased in the latter; and (iv) low expression throughout spermatogenesis (**Figure 3B-C**). We find that the majority of RMPs, including those involved in placing, removing m6A (VIRMA, YTHDC2, YTHDF2, ALKBH5, METTL14, METTL3) are highly expressed in spermatogonial cells, whereas their expression rapidly drops as the spermatogenic process begins (**Figure 3B-C**, see also **Figure S3C-D**).

Interestingly, this is not the case for all RMPs, such as m⁵C methyltransferase NSUN7, which is not expressed in early stages of spermatogenesis, but whose expression levels are drastically increased in spermatocytes and spermatids (**Figure S3A, 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 [55,56], 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.

We then investigated whether specific RMPs also showed increased expression patterns in epididymis, relative to other tissues (**Figure S3D**). Our analysis identified two RMPs as epididymis-enriched: (i) TRDMT1 -also known as DNMT2-, a N5-methyladenosine (m⁵C) methyltransferase modifying position 38 in specific tRNAs [57], and (ii) METTL1, a N7-methylguanosine (m⁷G) tRNA methyltransferase, which has been recently shown to act not only on tRNAs, but also on miRNAs, promoting their maturation [58]. Interestingly, TRDMT1 has been shown to play a major role in the transmission of paternal epigenetic information across generations [53]; 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 [59]. 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 m⁵C methyltransferase that has been shown to affect sperm motility [56,60]; (ii) NSUN2, an m⁵C tRNA methyltransferase involved in sperm differentiation [61]; (iii) METTL14, a component of the m⁶A 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] (**Figure 3D**, see also **Figure S3E**).

We found that NSUN7 is most highly expressed in chromatoid bodies of 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 [56,60,62,63] (**Figure 3D**, left panels). Intriguingly, NSUN7 displayed vesicular subcellular localisation 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 epididymis tissues. On the other hand, NSUN2 displayed high expression levels in the nucleolus of the spermatogonial cells and spermatocytes (**Figure 3D**, middle panels), supporting its role in sperm differentiation [61]. We also observed that NSUN2 is highly expressed in the initial segment of the epididymis, with decreased expression in the remaining epididymal sections.

METTL14 was also found highly expressed in early spermatogenesis and down-regulated during the subsequent stages (**Figure 3D**, right panels), in agreement with the dynamic regulation of m⁶A levels during spermatogenesis [34]. Finally, HENMT1 is mainly expressed in round and elongating spermatocytes, where it is accumulated in the chromatoid bodies (**Figure S3E**). Chromatoid bodies are known to contain large amounts of piRNAs, in agreement with the role that HENMT1 plays in piRNA stability during spermatogenesis [5]. 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 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 [64–66]. Several studies have shown that modulation of the RNA modification machinery can decrease the expression of specific oncogenes [36,67]. For example, in the case of glioblastoma, treatment with an FTO inhibitor was shown to decrease the expression levels of certain oncogenes [65]. Similarly, tRNA modifying enzymes NSUN2 and METTL1 can affect chemotherapy sensitivity by changing the methylation states of certain tRNAs [68]. Thus, understanding which epitranscriptomic players are dysregulated in each tumour type is essential to guide the research for future anticancer therapies targeting this regulatory layer.

To this end, we performed an integrative analysis of RMPs gene expression across 13,358 tumor-normal paired human samples gathered from publicly available datasets [69], which included 28 different cancer types (**Table S5**). Firstly, we compared the expression patterns between paired tumor-normal samples by measuring the log₂ fold changes of median gene expression between tumor and normal paired samples, for each RMP and cancer type (**Figure S4** 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 (**Figure S4**). Surprised by this result, we wondered whether these global up/down-regulation patterns could in fact be an artefact 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 (**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) ($r^2 = 0.98$) (**Figure S6**). 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 (**Figure 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 **Table S6**). Moreover, we find that the 'global' up/down-regulation patterns found using log₂ fold change comparisons are not further observed, suggesting that these results were in fact artefacts 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 down-regulated 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 up-regulated in two or more cancer types, as well as 8 RMPs which were consistently down-regulated in at least 2 cancer types (**Figure 4C**, see also **Table 1**). We found that the most frequently up-regulated RMP was HENMT1 (**Figure 4D**), a piRNA 2'-O-methyltransferase which is highly expressed in gonadal cells, involved in transposable element (TE) mutagenesis protection [5,70,71]. Whether the global up-regulation of HENMT1 in cancer samples might be contributing to increased TE mutagenesis is currently unknown.

The second most frequently up-regulated RMP was the L antigen family member 3 (LAGE3), a component of complex responsible for formation of N⁶-threonylcarbamoyladenosine (t⁶A) in position 37 of tRNAs (**Figure 4D**). Interestingly, this modification is found in the anticodon stem-loop of many tRNAs decoding ANN codons [72], and has been shown to affect both translation accuracy as well as efficiency[73].

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 (**Figure 5A**, see also **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 [74]. Similarly, our work revealed BUD23 expression to be correlated with cancer survival, in agreement with another recent study [75].

Surprisingly, we found that FTO expression levels are not significantly correlated with patient survival in LAML, despite studies are currently focusing on this cancer type to test FTO inhibitors (**Table S7**). By contrast, LAGE3 expression levels were significantly correlated with patient survival in LAML (**Figure 5B**). 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) (**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 (**Figure 1**), and find that duplications are often accompanied by the acquisition of restricted tissue expression patterns and/or change in its RNA target specificity. We find that the majority of tissue-restricted RMPs are in fact testis-enriched (**Figure 2**), suggesting that certain RMPs might play a pivotal role in sperm formation and maturation. Indeed, deletion of testis-enriched genes such as NSUN7, ADAD1 or HENMT1 leads to male sterility [5,55,56,60]

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 [49–52]. 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 [76]. In this regard, previous work has shown that METTL3/METTL14 mediated m⁶A modification is dynamically regulated in spermatogenesis [34]. Similarly, piRNA molecules in germ line 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 (**Figure 3**).

Recent works shown that specific RNA modifications are essential for the transmission of paternal diet-induced phenotypes intergenerationally [53]. Here we identify two RMPs (TRDMT1 and METTL1) whose expression is significantly enriched in epididymis (**Figure S3**), one of which (TRDMT1) was recently shown to be involved in the transmission of diet-induced paternal phenotypes across generations [53]. Whether METTL1 plays a role in intergenerational inheritance is yet to be deciphered; however, recent insights showing its role in miRNA maturation [58] suggest 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 [36,38,64,66,77], mostly focused on the machinery responsible for writing and erasing m⁶A modifications. For many years, FTO was thought to be of special interest due to its association with obesity [78]. However, later studies proved this genome-wide association to be false [79], and that the

single nucleotide variant present in the FTO intron was in fact associated with the activity of neighbouring genes [79].

Nonetheless FTO kept receiving special attention due to its perceived activity as an eraser of N6-methyladenosine (m6A) [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 (m6Am), which is much less abundant in mRNAs [28,80]. Similarly, FTO has been proposed to constitute a promising target for antitumor therapies [38,81,82]. While FTO has been shown to play an important role in leukaemia [81], 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 (**Figure 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 **Figure 4**). Moreover, we identify two enzymes, LAGE3 and HENMT1, as the top recurrently up-regulated 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 [83]; however, its role in tumorigenesis and cancer progression is yet to be determined. 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 m6A 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 in fact 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 (**Table S8**) by querying the PFAM database (<https://pfam.xfam.org/>). Each HMM profile then was used to query human proteome using `hmmsearch` function from HMMER software v.3.2.1 (<http://hmmer.org/>). Proteins above default threshold we kept as candidate RMW proteins (**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 [84]. Readers, erasers and non-catalytic subunit proteins were obtained from annotated Uniprot genes as well as from published literature [85]. 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 [86,87]

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) (**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 (**Table S11**). For each curated ortholog dataset, multiple sequence alignments were built using MAFFT with G-INS-1 method [88]. Alignment files were used to construct maximum-likelihood phylogenetic tree using IQ-Tree with bootstrapping (n=5000) [89]. Consensus trees were visualized using FigTree v 1.4.4 [90] and used to identify the duplication events (**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)[91]. 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, Transcripts Per Kilobase Million) were obtained from ENCODE3 (<http://genome.crg.es/~avlasova/ENCODE3/STAR-RSEM/>). 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.

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 [92]. 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 (orthologue genes) were plotted for the first two principal components using *ggplot* R package.

Analysis of RMPs expression during spermatogenesis

Processed spermatogenesis data [54] was extracted from GSE112393. 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 *complex heatmap* R package. Violin plots were built using the *ggplot* R package.

Immunohistochemistry

Testis and epididymis from 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 labelling 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). Brightfield images were acquired with a NanoZoomer-2.0 HT C9600 digital scanner (Hamamatsu) equipped with a 20X 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 (**Figure S6**).

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 [69]. 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 (up-regulated in tumor) and blue (down-regulated in tumor), and non-RMP proteins were depicted in grey.

Survival Analyses

Survival phenotypes were downloaded from the XENA Platform, using the “TCGA TARGET GTEX” cohort [69]. In order to analyse 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*.

DECLARATIONS

Code availability

All scripts used in this work will be made publicly available, and will be found at <https://github.com/obegik/RNAModMachinery>, once the paper is published.

Data availability

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) and raw IHC scans will be made publicly available when the paper is published.

Competing interests

The authors declare that the submitted work was carried out in the absence of any professional or financial relationship that could potentially be construed as a conflict of interest.

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.

Authors' contributions

OB performed all the analyses. 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. OB and EMN interpreted the results. EMN conceived the project. EMN and JSM supervised the work. OB and EMN wrote the manuscript, with input from all authors.

Acknowledgements

We thank all the members of the Mattick and Novoa labs for their valuable insights and discussion. The results shown here are in whole or part based upon data generated by the TCGA Research Network.

FIGURES

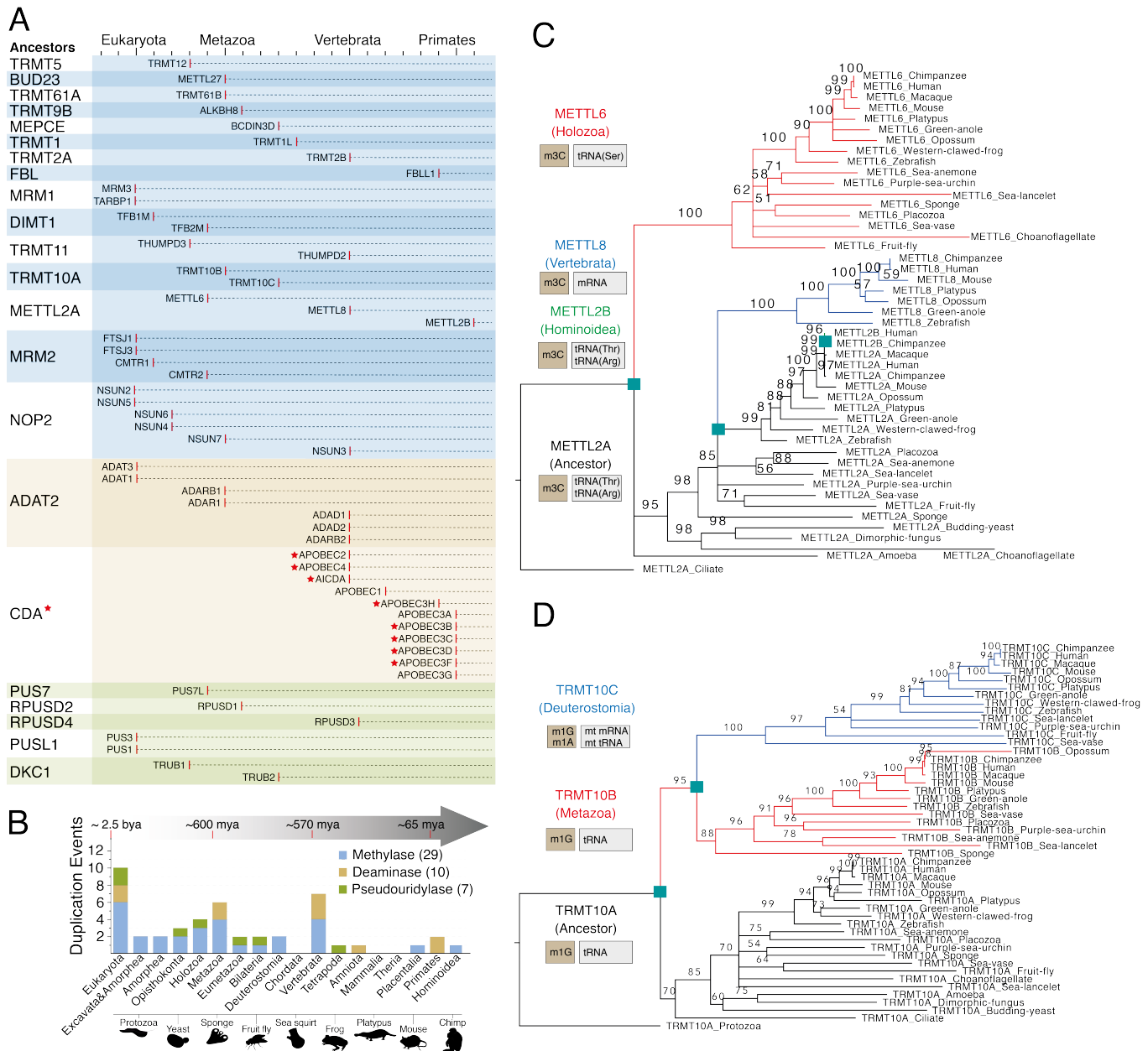


Figure 1. Evolutionary analysis of RNA modification ‘writers’. (A) Detailed overview of the evolutionary history of RMW duplications during eukaryotic evolution. Red stars indicate that proteins do not target RNAs but they are in the same family with an RNA writer protein. Red lines indicate the time where the enzyme has appeared. (B) Histogram of RMW duplication events throughout eukaryotic evolution. Duplication events were inferred by using multiple sequence alignments, coupled to maximum likelihood tree generation, for each family. (C and D) Maximum Likelihood phylogenetic trees of the methyltransferase family METTL2A/2B/6/8 (C) and TRMT10A/B/C (D). Cyan squares indicate the node where the duplication occurred. Numbers shown on the branches indicate bootstrapping values.

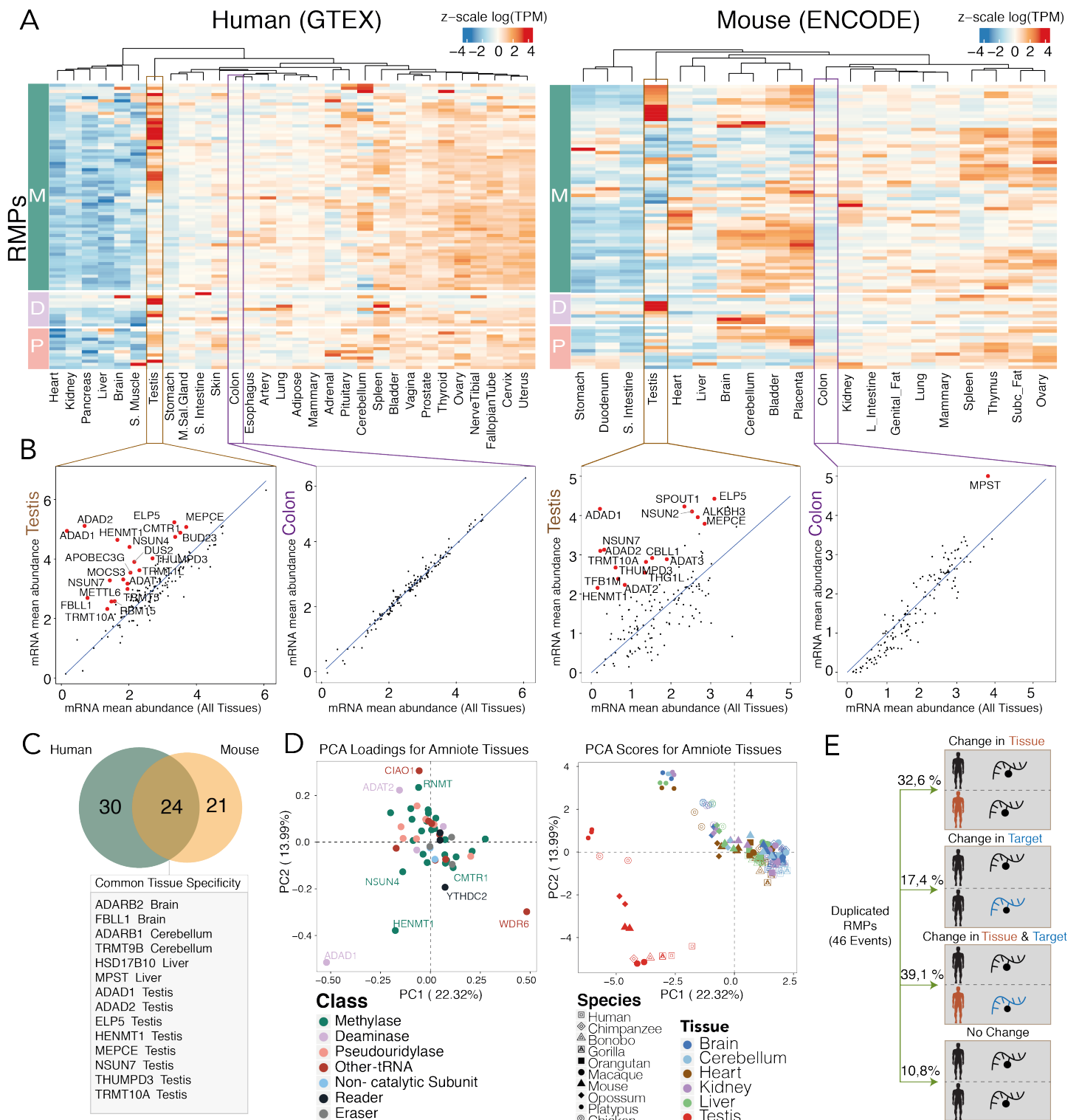


Figure 2. Analysis of RMP tissue specificity expression in different species. (A) Heatmap of z-scaled log(TPM) values of catalytic RNA writer proteins (M: methyltransferases; D: deaminases; P: pseudouridylases) throughout human and mouse tissues. In both human and mouse, testis has the most distinct RMP expression pattern in which many genes show really high expression, whereas other tissues such as colon shows 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 labelled in red. **(C)** Venn diagram of the conservation of tissue specificity between human and mouse. Out of 24 common tissue-specific genes, 14 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 scores 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 RMW duplication events shown in Figure 1, showing that 89% of them suffered a change in their tissue and/or target specificity.

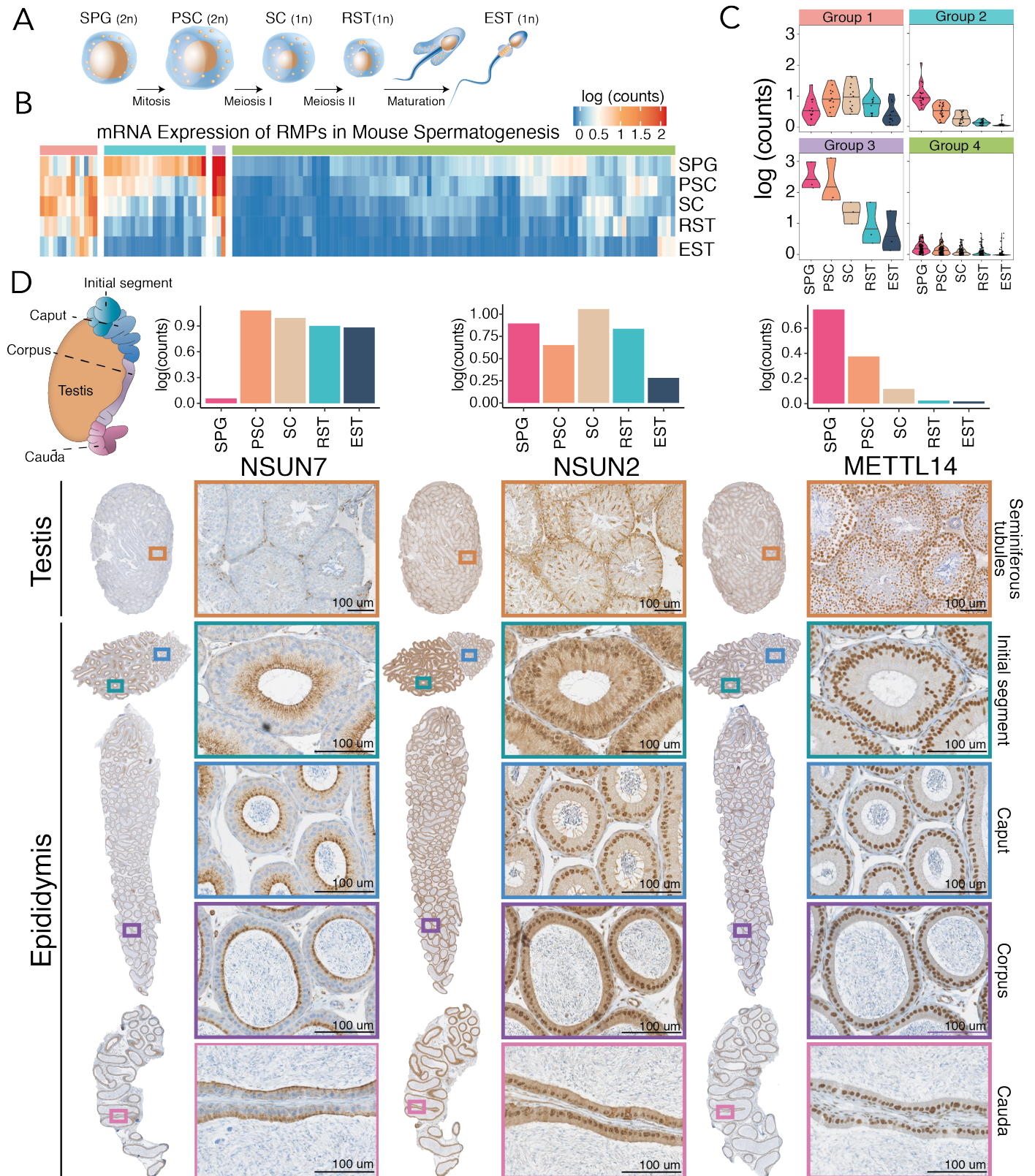


Figure 3. Analysis of RMP gene expression during spermatogenesis. (A) Schematic representation of the four main phases of spermatogenesis: i) mitotic division of spermatogonia (SPG) into primary spermatocytes (PSC) ii) meiotic division of PSCs into secondary spermatocytes (SC), iii) meiotic division SCs into round spermatids (RST) and iv) spermiogenesis, in which round spermatids (RST) mature into elongated spermatids (EST). (B) Heatmap of RMP expression levels in mouse testis. RMPs were clustered into 4 groups based on k-means analysis of their normalized average

mRNA expression values. **(C)** Violin plots of the expression patterns of each of the 4 identified clusters **(D)** RNA median expression barplot and immunohistochemistry of NSUN7, NSUN2 and METTL14, depicting distinct protein expression levels along the different sections of the testis and epididymis, as well as different subcellular localisations. Brown color indicates a specific staining of the antibody whereas blue represents hematoxylin counterstain.

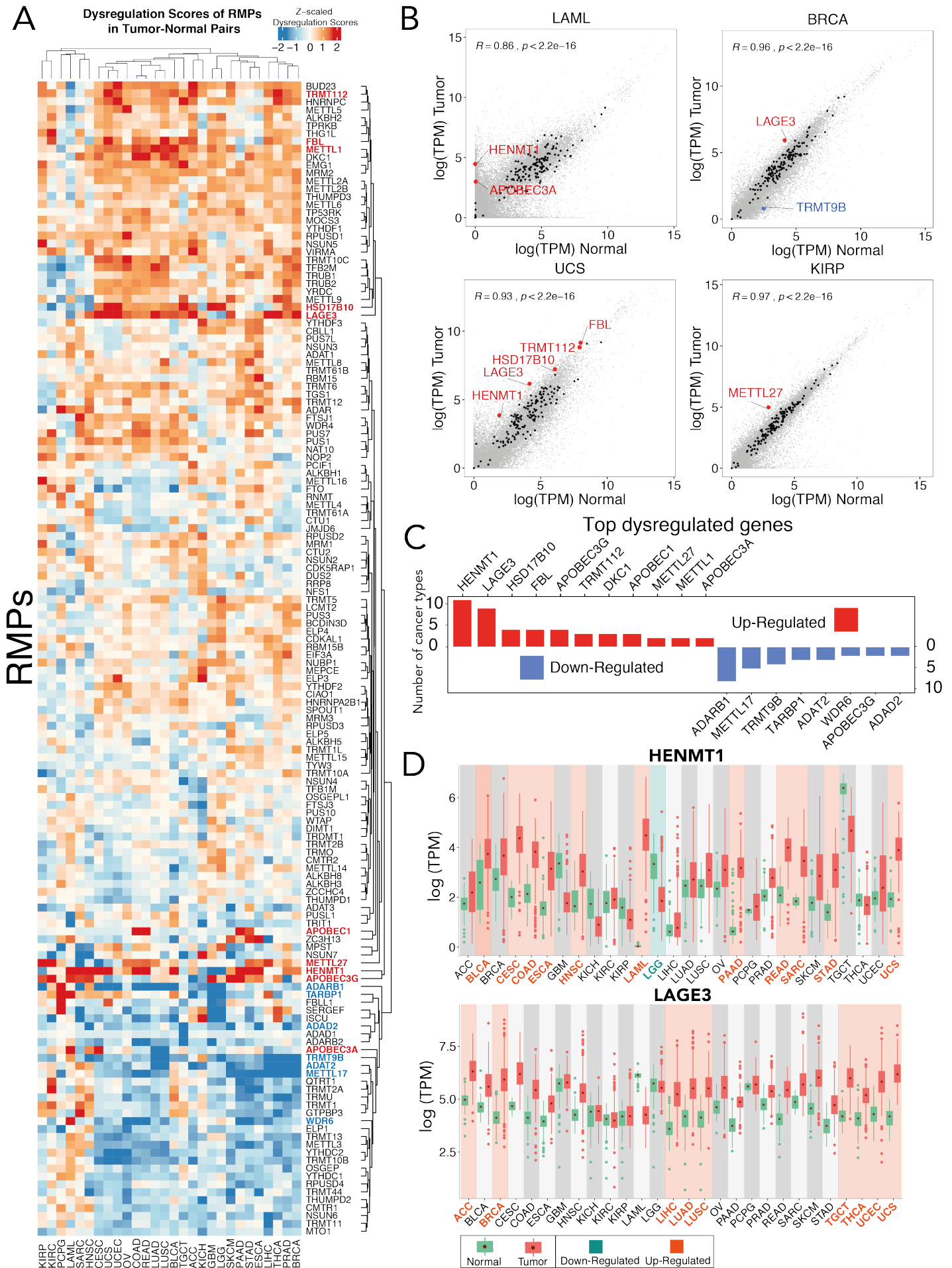


Figure 4. Expression analysis of RMPs in human tumor-normal paired samples. **(A)** Heatmap of z-scaled dysregulation scores of RMPs in tumor-normal paired samples, across 28 cancer types. Positive (red) values indicate up-regulation in tumor, whereas negative (blue) values indicate down-regulation. Genes labeled as red (up-regulated) and blue (down-regulated) 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 (up-regulated) or blue (down-regulated). Non-RMP genes are depicted in grey. **(C)** Barplot illustrating the number of cancer types in which significantly dysregulated genes highlighted in red (up-regulated) or blue (down-regulated). 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 down-regulated, whereas those highlighted in orange represent those cancer types in which the RMP is up-regulated. 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), PCPG (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).

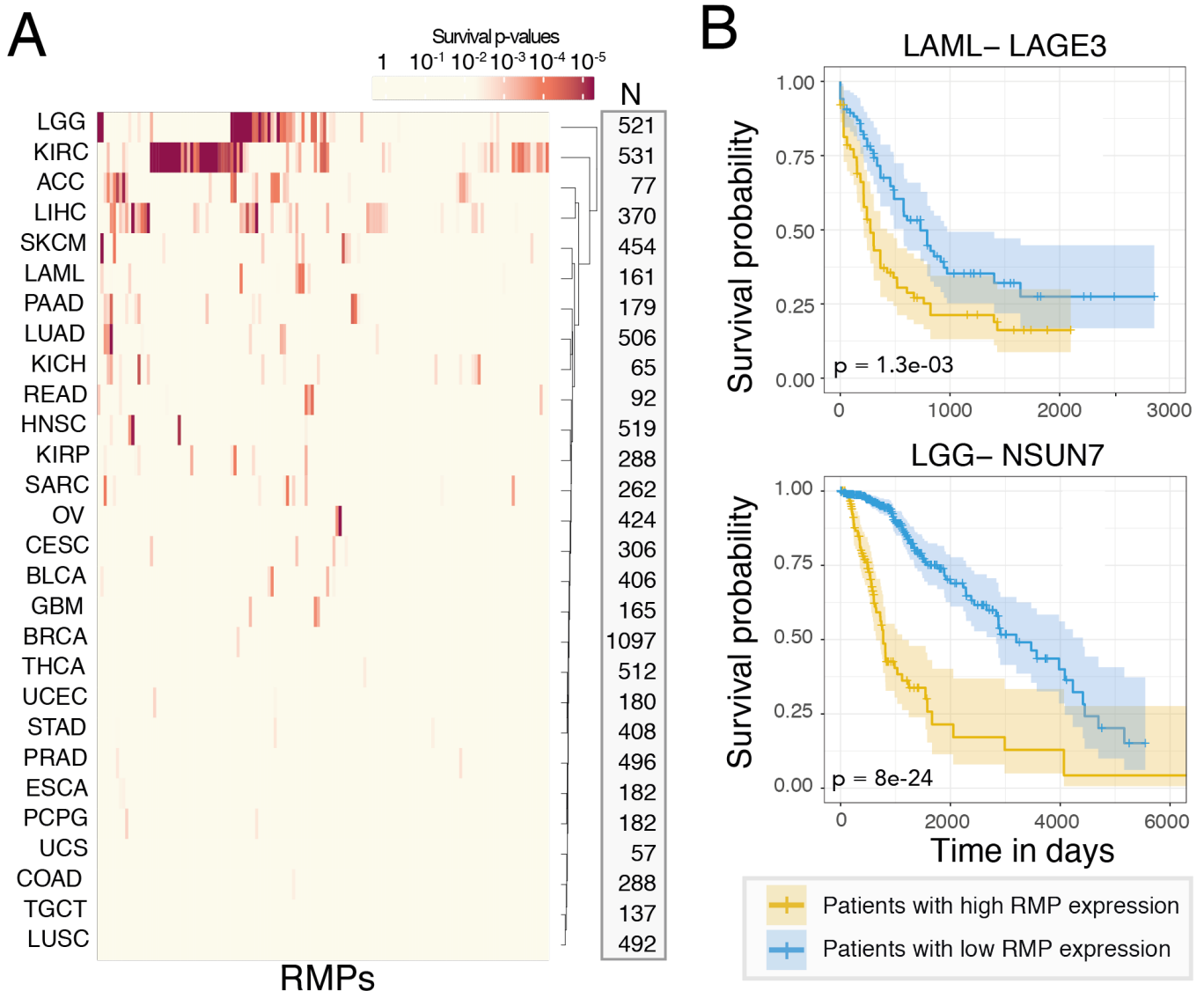


Figure 5. Prognostic value of RMPs expression levels in different cancer types. (A) Heatmap of survival p-values of 146 RMPs across 28 cancer types. Survival p-values are calculated by comparing the prognosis of patients that express high (upper 50%) versus low (lower 50%) RMP levels. “N” column shows the number of patients included for the analysis of each cancer type. (B) 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%).

TABLES

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

Short Name	Cancer type	Up-regulated RMPs	Down-regulated 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	

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* [Internet]. 2018;19. Available from: <http://dx.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 [Internet].* 2016; Available from: <https://www.ncbi.nlm.nih.gov/pubmed/28002401>

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 [Internet].* 2018; Available from: <http://dx.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 [Internet]. *Protein Science.* 2018. p. 233–44. Available from: <http://dx.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. 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.

50. Bettgowda A, Wilkinson MF. Transcription and post-transcriptional regulation of spermatogenesis. *Philos Trans R Soc Lond B Biol Sci.* 2010;365:1637–51.

51. Robles V, Herráez P, Labbé C, Cabrita E, Pšenička M, Valcarce DG, et al. Molecular basis of spermatogenesis and sperm quality. *Gen Comp Endocrinol.* 2017;245:5–9.

52. 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.

53. 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.

54. 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.

55. Connolly CM, Dearth AT, Braun RE. Disruption of murine *Tenr* results in teratospermia and male infertility. *Dev Biol.* 2005;278:13–21.

56. 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.

57. 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.
58. 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.
59. Vogel C, Marcotte EM. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat Rev Genet.* 2012;13:227–32.
60. 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.
61. 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.
62. 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.
63. Wolfson B, Gambone J, Rajfer J. Identification of motile sperm in caput epididymis. Intraoperative observations and clinical correlations. *Urology.* 1992;40:335–8.
64. Delaunay S, Frye M. RNA modifications regulating cell fate in cancer. *Nat Cell Biol.* 2019;21:552–9.
65. 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.
66. 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* [Internet]. 2019; Available from: <http://dx.doi.org/10.1016/j.stem.2019.03.021>
67. 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.
68. 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.
69. 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*, 326470. 2019;
70. 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.
71. Kirino Y, Mourelatos Z. 2'-O-methyl modification in mouse piRNAs and its methylase. *Nucleic Acids Symp Ser . Narnia;* 2007;51:417–8.
72. Thiaville PC, El Yacoubi B, Köhrer C. Essentiality of threonylcarbamoyladenine (t6A), a universal tRNA modification, in bacteria. *Molecular* [Internet]. Wiley Online Library; 2015; Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1111/mmi.13209>
73. 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.
74. 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* [Internet]. 2019; Available from:

<http://dx.doi.org/10.1007/s00401-019-02062-4>

75. Őunap K, Kasper L, Kurg A, Kurg R. The human WBSCR22 protein is involved in the biogenesis of the 40S ribosomal subunits in mammalian cells. *PLoS One*. 2013;8:e75686.
76. Roundtree IA, Evans ME, Pan T, He C. Dynamic RNA Modifications in Gene Expression Regulation. *Cell*. 2017;169:1187–200.
77. 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.
78. Fawcett KA, Barroso I. The genetics of obesity: FTO leads the way. *Trends Genet*. 2010;26:266–74.
79. 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.
80. 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.
81. 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 [Internet]. *Cancer Cell*. 2017. p. 127–41. Available from: <http://dx.doi.org/10.1016/j.ccell.2016.11.017>
82. 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.
83. 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.
84. de Crecy-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* [Internet]. 2019; Available from: <http://dx.doi.org/10.1093/nar/gkz011>
85. Lobo J, Costa AL, Cantante M, Guimaraes 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.
86. 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.
87. 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.
88. Katoh K, Rozewicki J, Yamada KD. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief Bioinform* [Internet]. 2017; Available from: <http://dx.doi.org/10.1093/bib/bbx108>
89. 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.
90. Rambaut A. FigTree v1. 4. 2012.
91. Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Proteomics. Tissue-based map of the human proteome. *Science*. 2015;347:1260419.
92. Brawand D, Soumillon M, Necsulea A, Julien P, Csardi G, Harrigan P, et al. The evolution of gene expression levels in mammalian organs. *Nature*. 2011;478:343–8.