ARTICLE

doi:10.1038/nature11112

Topology of the human and mouse m^{6}A RNA methylomes revealed by m^{6}A-seq

Dan Dominissini[1,2]*, Sharon Moshitch-Moshkovitz[1]*, Schraga Schwartz[3]*†, Mali Salmon-Divon[1], Lior Ungar[2,4], Sivan Osenberg[1,2], Karen Cesarkas[1], Jasmine Jacob-Hirsch[1], Ninette Amariglio[1], Martin Kupiec[4], Rotem Sorek[3] & Gideon Rechavi[1,2]

An extensive repertoire of modifications is known to underlie the versatile coding, structural and catalytic functions of RNA, but it remains largely uncharted territory. Although biochemical studies indicate that N^6-methyladenosine (m^{6}A) is the most prevalent internal modification in messenger RNA, an in-depth study of its distribution and functions has been impeded by a lack of robust analytical methods. Here we present the human and mouse m^{6}A modification landscape in a transcriptome-wide manner, using a novel approach, m^{6}A-seq, based on antibody-mediated capture and massively parallel sequencing. We identify over 12,000 m^{6}A sites characterized by a typical consensus in the transcripts of more than 7,000 human genes. Sites preferentially appear in two distinct landmarks—around stop codons and within long internal exons—and are highly conserved between human and mouse. Although most sites are well preserved across normal and cancerous tissues and in response to various stimuli, a subset of stimulus-dependent, dynamically modulated sites is identified. Silencing the m^{6}A methyltransferase significantly affects gene expression and alternative splicing patterns, resulting in modulation of the *p53* (also known as *TP53*) signalling pathway and apoptosis. Our findings therefore suggest that RNA decoration by m^{6}A has a fundamental role in regulation of gene expression.

Although the multiple layers of epigenetic regulation that result from modification of DNA and proteins have been intensively explored, RNA modifications are still uncharted territory[1]. The complex structure–function relationship of RNA makes it challenging to decipher its intricate biological roles. A wide variety of post-transcriptional modifications, with over a hundred known so far[2], decorate RNA molecules from all domains of life to expand their nucleotide repertoire. Nonetheless, knowledge of their location and function is limited at present.

m^{6}A is the most common internal messenger RNA modification found in eukaryotes, as well as in RNA of nuclear-replicating viruses[3]. It is catalysed by an evolutionarily conserved, nuclear, multi-component enzyme, only one of whose subunits, methyltransferase like 3 (*METTL3*), has been identified[4]. In all organisms tested, induced experimental deficiency of the methyltransferase is detrimental and leads to apoptosis (*Homo sapiens*[3]), developmental arrest (*Arabidopsis thaliana*[5]) or defects in gametogenesis (*Saccharomyces cerevisiae*[6], *Drosophila melanogaster*[7]).

Since m^{6}A was first discovered in mRNA decades ago[8], only a few sites have been mapped in cellular and viral RNA, none of them in human, both in coding and non-coding regions[9,10]. All sites were found within sequences conforming to the degenerate consensus RRACH (A = m^{6}A)[11,12]. Unlike adenosine-to-inosine editing, m^{6}A does not alter the coding capacity of transcripts[13,14]. Importantly, it was shown that a specific position can be methylated in only a fraction of transcripts[9] (that is, non-stoichiometric) and that certain transcripts seem to be completely devoid of m^{6}A (ref. 15), suggesting that this base modification serves some regulatory role. The recent identification of fat mass and obesity-associated protein (*FTO*) as the m^{6}A demethylase[16], and its dynamic association with nuclear speckles further supports this notion.

Progress towards elucidating the biological function of m^{6}A was largely impeded by a lack of efficient methods for its detection and manipulation[13]. Although clearly attesting to the essential role of the m^{6}A methyltransferase, previous studies provided only limited insight into the underlying molecular mechanisms by which this ubiquitous modification acts (reviewed in ref. 3).

To gain further insight into the role of m^{6}A in RNA metabolism, we sought to determine the positions of m^{6}As at a transcriptome-wide level. We present a novel approach, m^{6}A-seq, which accomplishes this goal by combining the high specificity of an anti-m^{6}A antibody applied to randomly fragmented transcripts with the power of massively parallel sequencing. Using m^{6}A-seq, we compile the first human and mouse RNA methylomes, to our knowledge, and demonstrate their evolutionary conservation and response to changing cellular conditions. We complement this analysis by characterizing the effects of m^{6}A on splicing, and by identifying protein candidates associated with methylated RNA sequences (Supplementary Fig. 1).

m^{6}A-seq exposes the RNA methylome

To identify and localize m^{6}A sites at a transcriptome-wide level we applied m^{6}A-seq to RNA purified from a human hepatocellular carcinoma cell line (HepG2). Poly(A)$^+$-selected RNA was fragmented into ~100-nucleotide-long oligonucleotides (input) and immunoprecipitated using an anti-m^{6}A affinity purified antibody. Libraries were prepared from immunoprecipitated as well as input control fragments, and subjected to massively parallel sequencing (Fig. 1a). Reads were uniquely aligned to a reference transcriptome containing a single, intron-free splice variant per gene (Supplementary Table 2), and m^{6}A sites were identified using a peak-detection algorithm with an estimated false detection rate (FDR) < 7% (Methods). This analysis yielded 12,769 putative m^{6}A sites (henceforth referred to as m^{6}A peaks) within 6,990 coding gene transcripts (Fig. 2e and Supplementary Fig. 2a–c) and 250 non-coding ones (Supplementary Fig. 2d),

[1]Cancer Research Center, Chaim Sheba Medical Center, Tel Hashomer 52621, Israel. [2]Sackler School of Medicine, Tel Aviv University, Tel Aviv 69978, Israel. [3]Department of Molecular Genetics, Weizmann Institute of Science, Rehovot 76100, Israel. [4]Department of Molecular Microbiology and Biotechnology, Tel Aviv University, Tel Aviv 69978, Israel. †Present address: Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA.
*These authors contributed equally to this work.

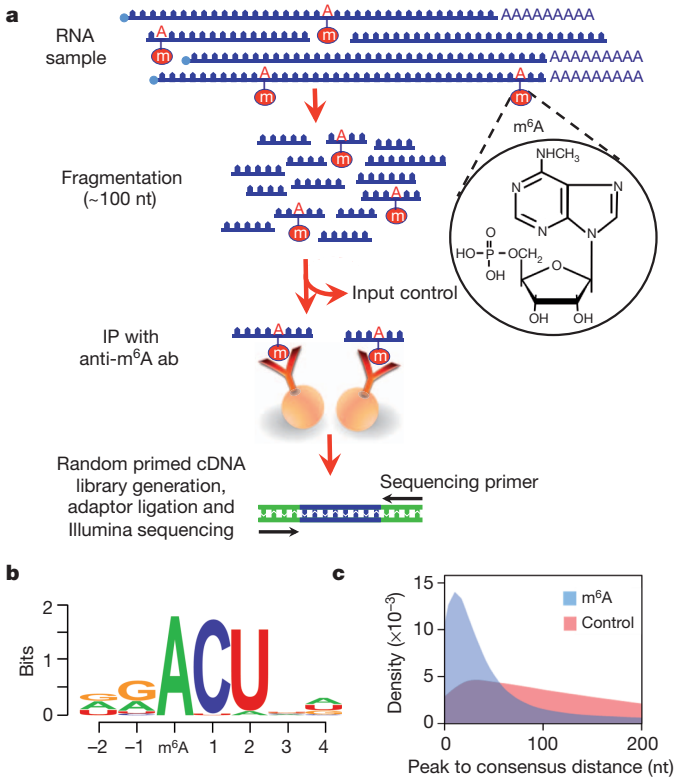

Figure 1 | m^{6}A-seq capture of modified RNA fragments exposes an enriched motif. **a**, Schematic diagram of the m^{6}A-seq protocol. ab, antibody; nt, nucleotide. IP, immunoprecipitation. **b**, Sequence logo representing the deduced consensus motif following clustering of all enriched motifs. **c**, Density plot illustrating the distribution of distance between the peaks of m^{6}A/control fragments and the nearest m^{6}A consensus motif (RRACU).

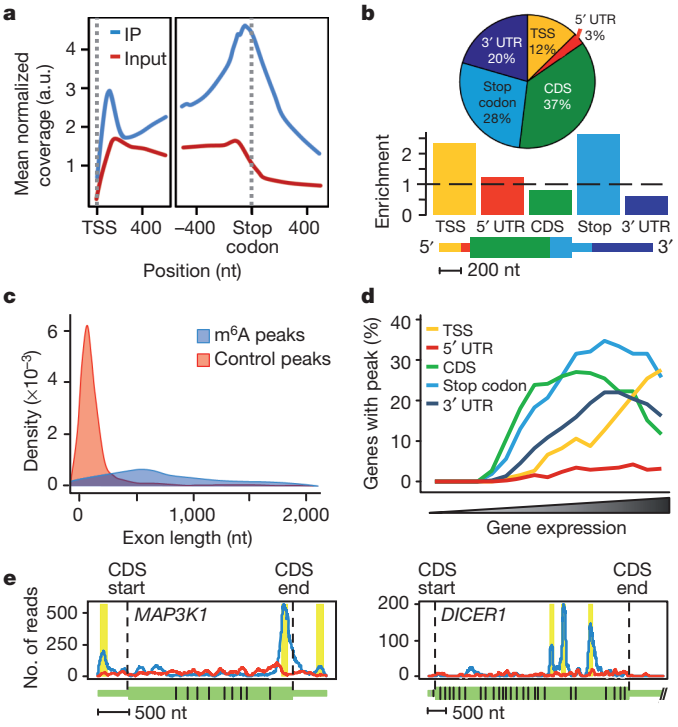

Figure 2 | The transcriptome landscape of m^{6}A reveals a unique topology. **a**, Metagene profiles depicting sequence coverage in windows surrounding the TSS (left) and stop codon (right). Coverage of m^{6}A IP and control (input) fragments is indicated in blue and red, respectively. To normalize for expression levels, coverage for each position is divided by the mean coverage in the control experiment. **b**, Top, pie chart presenting the fraction of m^{6}A peaks in each of five non-overlapping transcript segments. Middle, relative enrichment of m^{6}A peaks across transcript segments. Bottom, schematic of the five segments. **c**, Density plot showing the distribution of m^{6}A/control peaks according to exon length. **d**, Fraction of genes with m^{6}A peaks in each of the segments as a function of expression level. **e**, Representative gene transcripts harbouring m^{6}A peaks. Colour codes are the same as in panel **a**. Peaks are highlighted in yellow. Black dashed lines signify CDS borders; transcript architecture is shown beneath, with thin parts corresponding to UTRs and thicker ones to CDS; exon–exon junctions are indicated by vertical black lines. a.u., arbitrary units.

such as lincRNAs and antisense RNAs—comprising the first m^{6}A RNA methylome (Supplementary Tables 1, 6 and Supplementary Link 1). This number probably underestimates the actual number of m^{6}A sites (Supplementary Fig. 3).

Repeated analysis using the model-based analysis of ChIP-seq (MACS) peak-calling algorithm based on genome-aligned reads identified 20,401 peaks (FDR ≤ 5%, fold change ≥ 4), containing over 96% of the transcriptome-based peaks, among them 1,508 intron peaks (Supplementary Table 3).

The high specificity and immunoprecipitation compatibility of the antibody used are well established and have been described in a multitude of publications[9,16–22]. Several additional measures were taken to ensure the validity and stringency of our experimental approach (Supplementary Note 1), in addition to identification of a known m^{6}A site within 18S rRNA[23].

To determine whether our identified m^{6}A peaks share a common sequence element potentially specifying methylation, we performed an unbiased search for motifs enriched in regions surrounding m^{6}A peaks (Methods). Clustering of all significantly enriched sequences perfectly recapitulated the previously established RRAC core, and revealed a terminal U to be a dominant part of the consensus (Fig. 1b and Supplementary Fig. 4). The median distance between m^{6}A peaks and the consensus sequence was 24 nucleotides, compared to 117 nucleotides in control peaks, supporting the specificity of m^{6}A-seq and reflecting its resolution (Fig. 1c). The identification of a strong consensus reinforces the authenticity of the newly discovered m^{6}A peaks, and supports the existence of a predominant methylation machinery.

Our results show that the HepG2 transcriptome contains ~1 m^{6}A peak per 2,000 nucleotides (Supplementary Fig. 3), or ~1.7 peaks per gene. These results are in line with previous estimations of ~3 m^{6}A residues per average mRNA transcript in mammalian cells, which largely relied on the gross proportion of m^{6}As out of all methylated nucleosides[24]. Of note, as transcripts are fragmented before immunoprecipitation, the m^{6}A-seq analysis is blind to methylation patterns at the single transcript level, but rather produces a gene-specific methylation profile based on fragments derived from the entire transcript population of a given gene.

Given that the determined consensus is far more prevalent than the actual observed frequency of m^{6}As by an order of magnitude, other structural elements, sequence determinants and/or processing steps might be necessary to specify a site for methylation. A previous small-scale computational prediction suggested that m^{6}As lie within the loop of a stem–loop structure[3]. However, using RNAfold, we were unable to identify strong secondary structures in regions harbouring newly identified m^{6}A peaks.

As can be expected from such widespread occurrence, no functional category (GOTERM_MF_1-5) was found to be enriched in a gene ontology (GO) analysis, suggesting that m^{6}A has a fundamental function.

m^{6}A distribution reveals a unique topology

Having exposed a new layer of RNA modification, the next challenge is to infer putative functions. To this end, we began by systematically examining the distribution of m^{6}A peaks along transcripts relative to

landmarks in their architecture, based on the single-transcript, intron-free transcriptome data set (to avoid misinterpretations due to peaks falling into different isoform-dependent landmarks). We noted that m^{6}A peaks were markedly correlated with two distinct coordinates: immediately following the transcription start site (TSS) and in the vicinity of the stop codon (Fig. 2a). Overall, stop codon peaks were more pronounced than TSS peaks. To assess the enrichment methodically, we assigned each m^{6}A peak to one of five non-overlapping transcript segments: TSS; 5′ untranslated region (UTR); coding sequence (CDS); stop codon; and 3′ UTR (Fig. 2b, pie chart), and then normalized by the relative fraction each segment occupied in the transcriptome (Fig. 2b, histogram). The stop codon segment (400-nucleotide window centred on the stop codon) stood out as most enriched in m^{6}A peaks, with 28% of the peaks, representing a ~2.6-fold enrichment over the distribution expected by chance ($P < 2.2 \times 10^{-308}$, χ^2 test; Fig. 2e and Supplementary Fig. 5). Moreover, 33.5% of all adequately expressed genes (that is, >40 reads per kilobase (kb)) had a peak in this region. In only 2.3% of the cases did a methylation consensus form at the stop codon (incorporating its two consecutive purines), indicating that methylation typically did not occur precisely at the stop codon but rather in its vicinity.

Nonetheless, the much longer CDS segment, although slightly depleted in m^{6}A peaks when normalized by length, harboured the largest fraction of peaks (37%) (Fig. 2b, e). Notably, these peaks tend to occur within unusually long internal exons: of 2,838 m^{6}A peaks within internal exons, 2,453 (87%) are in exons longer than 400 nucleotides, compared to only 14% of the negative control peaks ($P = 3.4 \times 10^{-156}$, *t*-test) (Fig. 2c, Supplementary Fig. 6 and Supplementary Table 1). Comparing the lengths of flanking introns and the strengths of the 5′ and 3′ splice sites and polypyrimidine tract, we found only minor differences between long methylated exons and their unmethylated counterparts, which for the most part were not statistically significant.

In contrast to all other segments, we found no tendency for TSS peaks (that is, in the first 200 nucleotides of a transcript) to be near a methylation consensus sequence (Supplementary Fig. 7b), suggesting that a methyltransferase other than the *METTL3* complex is involved. Indeed, when adenosine is the first transcribed nucleotide, in addition to the obligatory ribose 2′-*O*-methylation, it can be further methylated by another methyltransferase at the N^6 position of the base to generate (N^6-,2′-*O*-)-dimethyladenosine (m^{6}A$_m$)[25,26], also recognizable by our antibody[21]. In support of this claim, transcripts with TSS peaks were ~25% more likely to begin with adenosine compared with transcripts lacking such peaks (Supplementary Fig. 7a). Taken together, TSS peaks appear to reflect, at least partly, m^{6}A$_m$ belonging to the 5′ cap structure, and attest, once more, to the validity of m^{6}A-seq.

We next examined the relationship between methylation and expression by plotting the fraction of genes with m^{6}A peaks in each of the segments as a function of expression level. Interestingly, the CDS, stop codon and 3′ UTR segments exhibited a non-monotonic relationship: whereas transcripts of moderately expressed genes were more likely to be methylated, transcripts of genes expressed at the two extremes were less methylated (Fig. 2d). This pattern, which is unlikely to be due to coverage limitations or to bias of differential read sampling (Supplementary Fig. 8), is of functional interest and is reminiscent of a similar relationship between DNA methylation found in gene bodies and their corresponding expression levels[27], raising the intriguing possibility that these phenomena might be connected. In contrast, we observed a positive correlation between expression level and the presence of an m^{6}A peak in the TSS segment.

Comparison of human and mouse methylomes

The non-random distribution of m^{6}A peaks in the human transcriptome indicates a fundamental role of this modification. To determine the evolutionary conservation and consequent functional importance of m^{6}A, the human and mouse methylomes were compared. We applied m^{6}A-seq to RNA purified from mouse liver and obtained 4,513 m^{6}A peaks within 3,376 coding gene transcripts and 66 non-coding ones (Supplementary Tables 1 and 7). Clustering of all significantly enriched sequences perfectly recapitulated the human methylation consensus sequence (Fig. 3a and Supplementary Fig. 9a). The mouse metagene profile revealed, as in human, a peak around the stop codon and at the TSS (Fig. 3b and Supplementary Fig. 9c, d). In mouse, too, modifications were highly enriched in long internal exons, with 91% of the peaks in exons longer than 400 nucleotides (Supplementary Table 1 and Supplementary Fig. 9b). Thus, on the global level, m^{6}A is highly conserved between the two species.

We next systematically assessed the extent of m^{6}A peak conservation on the gene level. Of the 4,513 identified mouse m^{6}A peaks, 2,023 could be reliably mapped to an adequately expressed, orthologous human position; 997 of them had an m^{6}A peak in the orthologous human position, representing 49% conservation. The highest extent of conservation is found at the stop codon segment, in which 57% of peaks are conserved, compared to 32–49% in other segments. Conservation was also higher in internal exons longer than 400 nucleotides than in shorter ones (56% and 44%, respectively).

The extent of conservation between human and mouse, which is highly significant over that which would be expected by chance ($P = 3.7 \times 10^{-136}$, Mann–Whitney test), is easily appreciated in the examples presented (Fig. 3c and Supplementary Fig. 10), and signifies that these sites are likely to be functional.

Methylation across conditions and cells

Accumulating evidence reveals that transfer RNA (tRNA) and ribosomal RNA (rRNA) modifications change in response to stimuli, suggesting a general model of dynamic control over RNA modification[28,29]. By analogy, we reasoned that participation of m^{6}A in active gene regulation may manifest itself as altered methylation profiles in response to changing cellular conditions or in a tissue-specific manner.

We put our hypothesis to the test by comparing the methylation profiles of untreated HepG2 cells to those of cells exposed to ultraviolet radiation, heat shock, hepatocyte growth factor (HGF; also known as scatter factor (SF)), and interferon-γ (Supplementary Table 2). Remarkably, all samples exhibited a marked similarity of

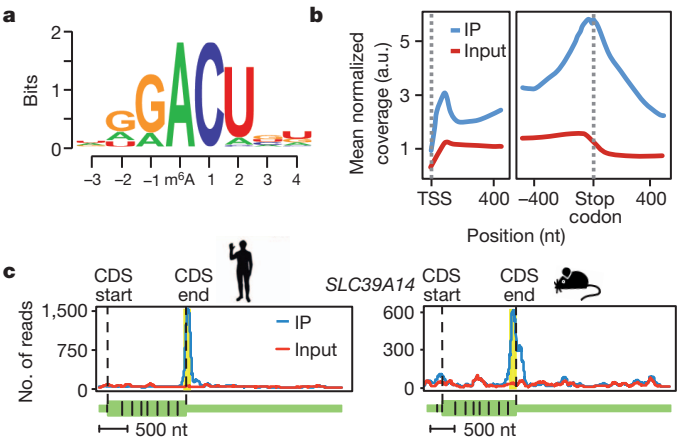

Figure 3 | m^{6}A methylome conservation between human and mouse. **a**, Sequence logo representing the deduced consensus motif following clustering of all enriched motifs. **b**, Metagene profiles depicting sequence coverage in windows surrounding the TSS (left) and stop codon (right). Colour codes are the same as in panel Fig. 2a. The mouse stop codon segment is even more enriched in m^{6}A peaks than its human counterpart, with 39% of all peaks located in this segment, corresponding to >3.5-fold enrichment over the distribution expected by chance ($P < 2.2 \times 10^{-308}$, χ^2 test). **c**, Orthologous genes with conserved m^{6}A peaks. Gene architecture is shown beneath. a.u., arbitrary units.

m^{6}A profiles, with 70–95% peak positions typically shared between conditions (Supplementary Fig. 11).

Nevertheless, we were able to detect a subset of treatment-dependent, dynamically altered peaks (Fig. 4 and Supplementary Table 8). These peaks did not correlate with absence of a proximal consensus motif. We note that our stringent approach probably underestimates the amount of differential m^{6}A peaks, all the more as it is insensitive to the proportion of methylated versions of a specific transcript.

We further examined tissue specificity and compared normal human brain to HepG2 cells. The former recapitulated all the global features of the m^{6}A methylome initially identified in HepG2 (including consensus motif and peak distribution along gene architecture) to demonstrate that the characteristics of the methylome are not an aberration of cancer tissues but rather a normal phenomenon (Supplementary Note 2).

m^{6}A affects RNA splicing

The function of m^{6}A in RNA metabolism was next assessed by silencing of *METTL3* in HepG2 cells, as its depletion was already shown to reduce the amount of m^{6}A in the transcriptome[4]. *METTL3* knockdown resulted in apoptosis (Supplementary Fig. 12), as expected[3]. Reads obtained from massively parallel sequencing of control (mock) and knockdown RNA were aligned to the genome and differential gene expression profiles were generated (Methods; Supplementary Table 4). Of 1,977 differentially expressed genes, 1,218 contain mapped m^{6}A peaks. Downregulated genes were significantly enriched with genes shown to have methylated introns ($P = 1.3 \times 10^{-7}$, hypergeometric test; Supplementary Fig. 13).

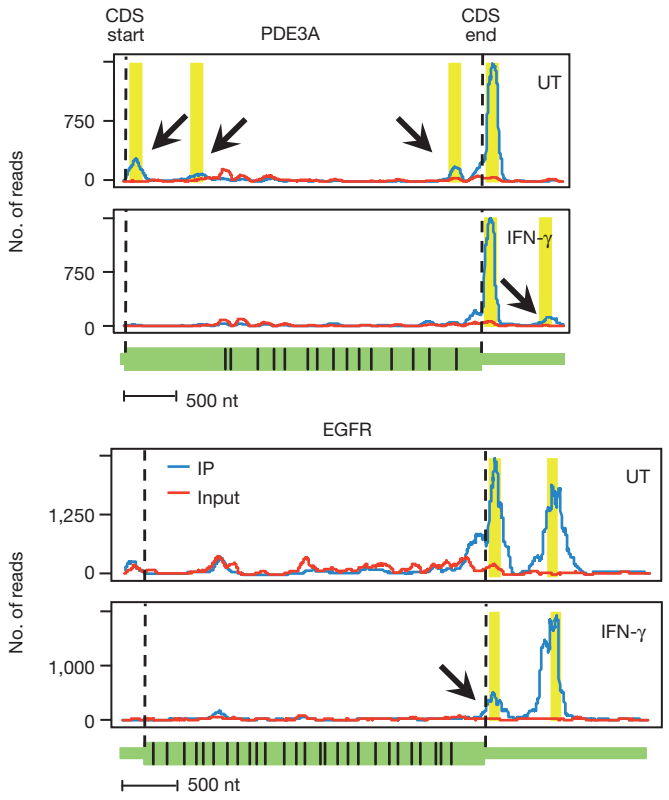

Figure 4 | Transcripts differing in methylation patterns under varying growing conditions. Arrows point at the differential position. The stability of the surrounding peaks underscores the authenticity of the changed peaks. Interferon-γ (IFN-γ) upregulates epidermal growth factor receptor (EGFR) RNA and protein levels[47,48], suggesting an association with absence of m^{6}A from the stop codon segment. PDE3A, phosphodiesterase 3A, cGMP inhibited; UT, untreated.

We reasoned that focusing on genes that were not differentially expressed (fold change < 2) but whose constituent isoforms were (fold change > 2), would help isolate a possible correlation between m^{6}A and isoform switching (between up- and downregulated isoforms). Indeed, methylated genes were overrepresented in this set (459/543, $P = 7.9 \times 10^{-12}$, hypergeometric test). Moreover, differentially spliced exons and introns were themselves significantly enriched with m^{6}A peaks: 99/474 exons ($P = 4 \times 10^{-15}$, hypergeometric test) and 811/2,672 introns ($P = 1 \times 10^{-105}$, hypergeometric test), further supporting a role for m^{6}A in splicing.

The salient findings revealed by *METTL3* knockdown prompted us to examine all m^{6}A peaks in the HepG2 transcriptome through the alternative splicing prism. Assigning peaks to either single- or multi-isoform coding genes revealed that the former are relatively undermethylated (6,489 and 11,946 peaks in 3,698 and 5,870 coding genes, respectively; $P = 5.6 \times 10^{-7}$, hypergeometric test). Accordingly, the average number of m^{6}A peaks per coding gene (1.93 peaks per gene) was higher in multi-isoform genes (2.04 peaks per gene) than in single-isoform ones (1.75 peaks per gene). Refining this observation further, assignment of all m^{6}A peaks to either constitutive or alternative spliced exons revealed that although 23.3% of coding exons are constitutive, they are significantly under-represented among all methylated sequences (10.9%, $P < 1 \times 10^{-305}$, hypergeometric test; Supplementary Fig. 14).

Interestingly, GO analysis of differentially expressed genes indicated a noteworthy enrichment of the *p53* signalling pathway (23 genes, corrected $P = 6.0 \times 10^{-5}$): 22/23 genes had differentially expressed splice variants, of which 18 were methylated. Moreover, 15 other members of the signalling pathway, which did not show significant differential expression at the gene level, exhibited significant differential expression at the isoform level (Supplementary Fig. 15a). For example, isoforms of *MDMX* (also known as *MDM4*), needed for p53 inactivation[30], were downregulated (Supplementary Fig. 15b). Similar pro-apoptotic effects were observed in other key genes belonging to this pathway (for example, *MDM2*, *FAS* and *BAX*). Modulation of *p53* signalling through splicing may be relevant to induction of apoptosis by silencing of *METTL3*.

Knockout of *IME4*, the yeast orthologue of *METTL3*, demonstrated its important role in regulation of the developmental switch from vegetative cells into gametogenesis, but failed to provide mechanistic insight (Supplementary Fig. 16).

Binding of RNA-interacting proteins

Methylation of RNA may also affect binding of interacting proteins, similar to recognition of 5-methylcytosine in DNA by specific binding proteins that mediate its repressive effects[31]. We observed an overlap between known RNA sequence elements and newly identified m^{6}A peaks: several internal ribosome entry sites (IRES)[32], the localization 'zip code' of β-actin[33] and the destabilization element of *c-Myc* (also known as *MYC*)[34], all harbour methylation peaks (Supplementary Fig. 2b, c and Supplementary Fig. 17).

An RNA affinity chromatography approach, using methylated and control versions of an RNA bait followed by mass spectrometry, was used to identify novel m^{6}A-binding proteins (Supplementary Fig. 18a).

Our analysis identified three RNA-binding proteins that may mediate novel connections between m^{6}A and cellular processes (Supplementary Fig. 18b and Supplementary Table 5). Two YTH (YT521-B homology) family proteins, *YTHDF2* and *YTHDF3*, bound exclusively to the methylated bait (Supplementary Fig. 18b, d). These proteins contain a recently characterized RNA-binding domain (YTH) that overlaps the methylated motif in our bait[35] (Supplementary Fig. 18c). Interestingly, the only two characterized YTH family members, the human *YT521-B* (also known as *YTHDC1*) and the fission yeast *Mmi1*, were implicated in alternative splice site selection and in specifying transcripts for nuclear degradation, respectively[36,37].

Another protein that was significantly associated with the m^{6}A bait was *ELAVL1* (also known as *HUR*)[38] (Supplementary Fig. 18b, d).

Discussion

Lack of knowledge regarding the distribution of m^{6}A in the transcriptome was one of the major factors that limited understanding of its role in RNA metabolism. m^{6}A-seq provides a valuable tool to bridge this gap and allows global mapping of the methylome to uncover several of its fundamental properties. The human and mouse RNA methylomes presented here reveal that RNA is non-randomly punctuated by m^{6}A throughout the body of most transcripts. We observe that m^{6}A sites are correlated with two distinct landmarks with telling biological importance: around stop codons and within long internal exons. The fact that this architectural pattern is also the most evolutionary conserved feature of the RNA methylome indicates an elementary role for this modification.

The unique distribution of m^{6}A provides hints as to its functions. It is unnecessary to assume one unifying principle behind this modification—m^{6}A can be involved in several cellular functions, similar to inosine, the best-studied global RNA modification[39]. The discovery of proteins that preferentially bind to methylated RNA raises the possibility that they are involved in mediating its biological effects.

Methylation of long internal exons is suggestive of involvement in splicing. Perhaps it reflects an auxiliary measure to alleviate the constraints imposed by larger exons on the exon-definition machinery[40,41]. The correlation of methylation with multi-isoform genes and differentially spliced exons and introns indicates a role for m^{6}A in splicing control. In line with this hypothesis, *METTL3* and *FTO* co-localize with splicing proteins in nuclear speckles[4,16], and the application of methylation inhibitors to cells resulted in nuclear accumulation of unspliced transcripts[42–44]. Integration of the methylation layer into 'splicing codes', such as that previously described[45], may improve the predictive power of alternative splicing events.

Methylation in the vicinity of stop codons hints at the direction of translational control. It is tempting to speculate that the presence of m^{6}A around this landmark affects translation efficiency, either directly or through the recruitment of specific factors. Indeed, the link between m^{6}A and translational efficiency was previously demonstrated *in vitro*[46].

The dynamic regulation of nucleic-acid post-transcriptional modifications, analogous to post-translational protein modifications, is just beginning to be elucidated[28]. Here we provide initial evidence for a change in m^{6}A positions and frequencies in response to various perturbations. Given the previously established non-stoichiometric nature of this base modification[9] and existence of the *FTO* demethylase[16], it seems reasonable to speculate a role for it as part of a global cellular response to stimuli.

The m^{6}A methylome opens new avenues for correlating the methylation layer with other processing levels. In many ways, this approach is a forerunner, providing a reference and paving the way for the uncovering of other RNA modifications, which together constitute a new realm of biological regulation, recently termed RNA epigenetics[1].

METHODS SUMMARY

RNA was randomly fragmented and subjected to immunoprecipitation using an anti-m^{6}A antibody. Immunoprecipitated and input control samples were sequenced using Illumina GAIIx. Using both in-house and MACS peak-calling algorithms, regions enriched in immunoprecipitated relative to input samples were identified as m^{6}A peaks comprising the methylome. Motifs enriched around m^{6}A peaks were identified. Human and mouse methylomes were compared by analysing methylation profiles of orthologous genes. *METTL3* knockdown was achieved by short interfering RNA (siRNA) transfection. Control and *METTL3* knockdown samples were sequenced using Illumina GAIIx and subjected to differential expression analyses. RNA affinity chromatography experiments were carried out using two biotinylated RNA baits differing in a single m^{6}A modification and bound proteins were analysed by mass spectrometry (LC-MS/MS).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 24 April 2011; accepted 11 April 2012.
Published online 29 April 2012.

1. He, C. Grand challenge commentary: RNA epigenetics? *Nature Chem. Biol.* **6,** 863–865 (2010).
2. Cantara, W. A. *et al.* The RNA Modification Database, RNAMDB: 2011 update. *Nucleic Acids Res.* **39,** D195–D201 (2011).
3. Bokar, J. A. in *Fine-Tuning of RNA Functions by Modification and Editing* Vol. 12 (ed. Grosjean, H.) 141–177 (Springer, 2005).
4. Bokar, J. A., Shambaugh, M. E., Polayes, D., Matera, A. G. & Rottman, F. M. Purification and cDNA cloning of the AdoMet-binding subunit of the human mRNA (N^6-adenosine)-methyltransferase. *RNA* **3,** 1233–1247 (1997).
5. Zhong, S. *et al.* MTA is an *Arabidopsis* messenger RNA adenosine methylase and interacts with a homolog of a sex-specific splicing factor. *Plant Cell* **20,** 1278–1288 (2008).
6. Clancy, M. J., Shambaugh, M. E., Timpte, C. S. & Bokar, J. A. Induction of sporulation in *Saccharomyces cerevisiae* leads to the formation of *N*^{6}-methyladenosine in mRNA: a potential mechanism for the activity of the IME4 gene. *Nucleic Acids Res.* **30,** 4509–4518 (2002).
7. Hongay, C. F. & Orr-Weaver, T. L. *Drosophila* Inducer of MEiosis 4 (IME4) is required for Notch signaling during oogenesis. *Proc. Natl Acad. Sci. USA* **108,** 14855–14860 (2011).
8. Desrosiers, R., Friderici, K. & Rottman, F. Identification of methylated nucleosides in messenger RNA from Novikoff hepatoma cells. *Proc. Natl Acad. Sci. USA* **71,** 3971–3975 (1974).
9. Horowitz, S., Horowitz, A., Nilsen, T. W., Munns, T. W. & Rottman, F. M. Mapping of N6-methyladenosine residues in bovine prolactin mRNA. *Proc. Natl Acad. Sci. USA* **81,** 5667–5671 (1984).
10. Kane, S. E. & Beemon, K. Precise localization of m6A in Rous sarcoma virus RNA reveals clustering of methylation sites: implications for RNA processing. *Mol. Cell. Biol.* **5,** 2298–2306 (1985).
11. Harper, J. E., Miceli, S. M., Roberts, R. J. & Manley, J. L. Sequence specificity of the human mRNA N6-adenosine methylase *in vitro*. *Nucleic Acids Res.* **18,** 5735–5741 (1990).
12. Wei, C. M. & Moss, B. Nucleotide sequences at the N^6-methyladenosine sites of HeLa cell messenger ribonucleic acid. *Biochemistry* **16,** 1672–1676 (1977).
13. Dai, Q. *et al.* Identification of recognition residues for ligation-based detection and quantitation of pseudouridine and *N*^{6}-methyladenosine. *Nucleic Acids Res.* **35,** 6322–6329 (2007).
14. Levanon, E. Y. *et al.* Systematic identification of abundant A-to-I editing sites in the human transcriptome. *Nature Biotechnol.* **22,** 1001–1005 (2004).
15. Perry, R. P. & Scherrer, K. The methylated constituents of globin mRNA. *FEBS Lett.* **57,** 73–78 (1975).
16. Jia, G. *et al.* *N*^{6}-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. *Nature Chem. Biol.* **7,** 885–887 (2011).
17. Bringmann, P. & Luhrmann, R. Antibodies specific for *N*^{6}-methyladenosine react with intact snRNPs U2 and U4/U6. *FEBS Lett.* **213,** 309–315 (1987).
18. Dante, R. & Niveleau, A. Inhibition of *in vitro* translation by antibodies directed against *N*^{6}-methyladenosine. *FEBS Lett.* **130,** 153–157 (1981).
19. Munns, T. W., Liszewski, M. K., Oberst, R. J. & Sims, H. F. Antibody nucleic acid complexes. Immunospecific retention of N6-methyladenosine-containing transfer ribonucleic acid. *Biochemistry* **17,** 2573–2578 (1978).
20. Munns, T. W., Liszewski, M. K. & Sims, H. F. Characterization of antibodies specific for N^6-methyladenosine and for 7-methylguanosine. *Biochemistry* **16,** 2163–2168 (1977).
21. Munns, T. W., Oberst, R. J., Sims, H. F. & Liszewski, M. K. Antibody-nucleic acid complexes. Immunospecific recognition of 7-methylguanine- and *N*^{6}-methyladenine-containing 5′-terminal oligonucleotides of mRNA. *J. Biol. Chem.* **254,** 4327–4330 (1979).
22. Munns, T. W., Sims, H. F. & Liszewski, M. K. Immunospecific retention of oligonucleotides possessing *N*^{6}-methyladenosine and 7-methylguanosine. *J. Biol. Chem.* **252,** 3102–3104 (1977).
23. Czerwoniec, A. *et al.* MODOMICS: a database of RNA modification pathways. 2008 update. *Nucleic Acids Res.* **37,** D118–D121 (2009).
24. Perry, R. P., Kelley, D. E., Friderici, K. & Rottman, F. The methylated constituents of L cell messenger RNA: evidence for an unusual cluster at the 5′ terminus. *Cell* **4,** 387–394 (1975).
25. Keith, J. M., Ensinger, M. J. & Mose, B. HeLa cell RNA (2′-*O*-methyladenosine-*N*^{6}-)-methyltransferase specific for the capped 5′-end of messenger RNA. *J. Biol. Chem.* **253,** 5033–5039 (1978).
26. Wei, C., Gershowitz, A. & Moss, B. *N*^{6}, *O*$^{2'}$-dimethyladenosine a novel methylated ribonucleoside next to the 5′ terminal of animal cell and virus mRNAs. *Nature* **257,** 251–253 (1975).
27. Zilberman, D., Gehring, M., Tran, R. K., Ballinger, T. & Henikoff, S. Genome-wide analysis of *Arabidopsis thaliana* DNA methylation uncovers an interdependence between methylation and transcription. *Nature Genet.* **39,** 61–69 (2007).
28. Chan, C. T. *et al.* A quantitative systems approach reveals dynamic control of tRNA modifications during cellular stress. *PLoS Genet.* **6,** e1001247 (2010).
29. Schaefer, M. *et al.* RNA methylation by Dnmt2 protects transfer RNAs against stress-induced cleavage. *Genes Dev.* **24,** 1590–1595 (2010).
30. Lenos, K. & Jochemsen, A. G. Functions of MDMX in the modulation of the p53-response. *J. Biomed. Biotechnol.* **2011,** 876173 (2011).

31. Klose, R. J. & Bird, A. P. Genomic DNA methylation: the mark and its mediators. *Trends Biochem. Sci.* **31,** 89–97 (2006).
32. Mokrejs, M. *et al.* IRESite—a tool for the examination of viral and cellular internal ribosome entry sites. *Nucleic Acids Res.* **38,** D131–D136 (2009).
33. Kislauskis, E. H., Zhu, X. & Singer, R. H. Sequences responsible for intracellular localization of β-actin messenger RNA also affect cell phenotype. *J. Cell Biol.* **127,** 441–451 (1994).
34. Bernstein, P. L., Herrick, D. J., Prokipcak, R. D. & Ross, J. Control of c-myc mRNA half-life *in vitro* by a protein capable of binding to a coding region stability determinant. *Genes Dev.* **6,** 642–654 (1992).
35. Zhang, Z. *et al.* The YTH domain is a novel RNA binding domain. *J. Biol. Chem.* **285,** 14701–14710 (2010).
36. Harigaya, Y. *et al.* Selective elimination of messenger RNA prevents an incidence of untimely meiosis. *Nature* **442,** 45–50 (2006).
37. Rafalska, I. *et al.* The intranuclear localization and function of YT521-B is regulated by tyrosine phosphorylation. *Hum. Mol. Genet.* **13,** 1535–1549 (2004).
38. Brennan, C. M. & Steitz, J. A. HuR and mRNA stability. *Cell. Mol. Life Sci.* **58,** 266–277 (2001).
39. Nishikura, K. Functions and regulation of RNA editing by ADAR deaminases. *Annu. Rev. Biochem.* **79,** 321–349 (2010).
40. Keren, H., Lev-Maor, G. & Ast, G. Alternative splicing and evolution: diversification, exon definition and function. *Nature Rev. Genet.* **11,** 345–355 (2010).
41. Sterner, D. A., Carlo, T. & Berget, S. M. Architectural limits on split genes. *Proc. Natl Acad. Sci. USA* **93,** 15081–15085 (1996).
42. Camper, S. A., Albers, R. J., Coward, J. K. & Rottman, F. M. Effect of undermethylation on mRNA cytoplasmic appearance and half-life. *Mol. Cell. Biol.* **4,** 538–543 (1984).
43. Carroll, S. M., Narayan, P. & Rottman, F. M. N6-methyladenosine residues in an intron-specific region of prolactin pre-mRNA. *Mol. Cell. Biol.* **10,** 4456–4465 (1990).
44. Stoltzfus, C. M. & Dane, R. W. Accumulation of spliced avian retrovirus mRNA is inhibited in *S*-adenosylmethionine-depleted chicken embryo fibroblasts. *J. Virol.* **42,** 918–931 (1982).
45. Barash, Y. *et al.* Deciphering the splicing code. *Nature* **465,** 53–59 (2010).
46. Tuck, M. T., Wiehl, P. E. & Pan, T. Inhibition of 6-methyladenine formation decreases the translation efficiency of dihydrofolate reductase transcripts. *Int. J. Biochem. Cell Biol.* **31,** 837–851 (1999).
47. Hamburger, A. W. & Pinnamaneni, G. Interferon-induced enhancement of transforming growth factor-α expression in a human breast cancer cell line. *Proc. Soc. Exp. Biol. Med.* **202,** 64–68 (1993).
48. Mujoo, K., Donato, N. J., Lapushin, R., Rosenblum, M. G. & Murray, J. L. Tumor necrosis factor α and γ-interferon enhancement of anti-epidermal growth factor receptor monoclonal antibody binding to human melanoma cells. *J. Immunother. Emphasis Tumor Immunol.* **13,** 166–174 (1993).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank H. Cedar (The Hebrew University, Jerusalem) for his comments. We thank the Kahn Family Foundation for their support. This work was supported in part by grants from the Flight Attendant Medical Research Institute (FAMRI), Bio-Med Morasha ISF (grant no. 1942/08), and The Israel Ministry for Science and Technology (Scientific Infrastructure Program). R.S. was supported by the ERC-StG program (grant 260432). G.R. holds the Djerassi Chair in Oncology at the Sackler Faculty of Medicine, Tel Aviv University. This work was performed in partial fulfilment of the requirements for a PhD degree to D.D., Sackler Faculty of Medicine, Tel Aviv University.

Author Contributions D.D., S.M.-M. and G.R. conceived and designed the experiments; D.D., S.M.-M., L.U., K.C., S.O. and J.J.-H. performed the experiments; S.S., M.S.-D. and R.S. performed the bioinformatic analysis; D.D., S.M.-M., M.S.-D., N.A., M.K., S.S., R.S. and G.R. analysed and interpreted results, and wrote the paper.

Author Information Data have been deposited in NCBI's Gene Expression Omnibus (GEO) and are accessible through GEO series accession number GSE37005 (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE37005). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to G.R. (Gidi.Rechavi@sheba.health.gov.il).

METHODS

Tissues, cell lines and strains. Human hepatocellular carcinoma cell line, HepG2, was maintained in DMEM (Gibco, Invitrogen) containing 4.5 g l^{-1} glucose and L-glutamine supplemented with 10% FBS and penicillin/streptomycin. Where indicated, HepG2 cells were incubated with IFN-γ (200 ng ml^{-1}) or HGF (10 ng ml^{-1}) overnight. Stress effects were tested in HepG2 cells by either 30 min incubation at 43 °C (heat shock) or ultraviolet irradiation of 0.04 J cm^{-2} at 254 nm. Cells were allowed to recover for 4 h in normal growing conditions before harvesting using trypsin.

Yeast. Wild-type and *ime4Δ* mutant (created by conventional knockout protocols) SK1 *S. cerevisiae* cells were grown vegetatively in YPD medium (yeast extract, peptone, dextrose), and transferred to sporulation medium (1% potassium acetate) for induction of sporulation for 4 h. RNA was purified from the cells using TRIzol (Invitrogen) and glass beads, and hybridized to Affymetrix microarrays. Experiments were conducted in biological replicates.

Microarrays. Biotinylated cDNA was prepared according to standard Affymetrix protocol from 600 ng total RNA. After fragmentation, 10 μg of cRNA were hybridized for 16 h at 45 °C on yeast 2.0 Genome Array. GeneChips were washed and stained in the Affymetrix Fluidics Station 400. Feature intensity values from scanned arrays were normalized and reduced to expression summaries using the RMA algorithm. Control probes and probe sets with an interquartile range below 0.1 were removed.

RNA preparation. Total human brain RNA was purchased from Biochain. Total RNA from HepG2 cell pellets was extracted using PerfectPure RNA cultured cell kit (5 Prime). Total RNA of C57/BL6 mouse livers was extracted using PerfectPure RNA tissue kit (5 Prime). To avoid DNA contaminations all samples were treated with DNase (5 Prime).

Enrichment of polyadenylated RNA (polyA$^+$ RNA) from total RNA was performed using one round of GenElute mRNA miniprep kit (Sigma-Aldrich).

RNA samples were chemically fragmented into ~100-nucleotide-long fragments by 5 min incubation at 94 °C in fragmentation buffer (10 mM $ZnCl_2$, 10 mM Tris-HCl, pH 7). The fragmentation reaction was stopped with 0.05 M EDTA, followed by standard ethanol precipitation. Samples were resuspended in H_2O at ~1 μg μl^{-1} concentration and subjected to m^{6}A-seq.

m^{6}A-seq. Fragmented RNA (400 μg mRNA or 2.5 mg total RNA) was incubated for 2 h at 4 °C with 5 μg of affinity purified anti-m^{6}A polyclonal antibody (Synaptic Systems) in IPP buffer (150 mM NaCl, 0.1% NP-40, 10 mM Tris-HCl, pH 7.4). The mixture was then immunoprecipitated by incubation with protein-A beads (Repligen) at 4 °C for an additional 2 h. After extensive washing, bound RNA was eluted from the beads with 0.5 mg ml^{-1} N^6-methyladenosine (Sigma-Aldrich) in IPP buffer, and ethanol precipitated. RNA was resuspended in H_2O and used for library generation with mRNA sequencing kit (Illumina). Sequencing was carried out on Illumina Genome Analyzer (GAIIx) according to the manufacturer's instructions, using a 10 pM template per sample for cluster generation, and genomic sequencing kit V2 (Illumina).

Gene silencing by siRNA. Scrambled control or siRNA directed against *METTL3* (Invitrogen, 5′-AUAUCACAACAGAUCCACUGAGGUG-3′) were transfected into HepG2 cells using oligofectamine reagent (Invitrogen). Cells were re-transfected after 48 and 96 h, and harvested on days 2, 5 and 7, to allow more efficient METTL3 silencing. RNA purification and massively parallel sequencing were performed as described above. The experiment was conducted in two biological replicates, generating ~55 and ~70 million reads in knockdown and control respectively.

TUNEL assay. METTL3 knockdown and mock control cells were assayed for apoptosis on day 7 using DeadEnd Fluorometric TUNEL System (Promega). Nuclei were stained with DAPI (Sigma-Aldrich). Samples were immediately visualized using LSM 510 confocal microscope (Zeiss).

RNA affinity chromatography, mass spectrometry and western blot analysis. Two biotin-labelled RNA oligonucleotide baits spanning 42 nucleotides centred on a previously characterized m^{6}A site in the RSV genome were synthesized, only one of which was N^6-methylated: 5′-biotin-AUGGGCCGUUCAUCUGCUAAAAGG-m^{6}A- CUGCUUUUGGGGCUUGU-3′ and 5′-biotin-AUGGGCCGUUCAUCUGCUAAAAGGACUGCUUUUGGGGCUUGU-3′. HepG2 cells were harvested at 70–80% confluence, washed with PBS and lysed in lysis buffer (10 mM NaCl, 2 mM EDTA,0.5% Triton X-100, 0.5 mM DTT, 10 mM Tris-HCl, pH 7.5) containing complete protease inhibitor cocktail (Roche) and phosphatase inhibitor cocktail 2 (Sigma-Aldrich). Lysates were separated from insoluble cell debris by centrifugation (10,000*g* for 15 min at 4 °C) and pre-cleared for 1 h at 4 °C by incubation with streptavidin-conjugated agarose beads (Sigma-Aldrich) in binding buffer (150 mM KCl, 1.5 mM $MgCl_2$, 0.05% (v/v) NP-40, 0.5 mM DTT, 10 mM Tris-HCl pH 7.5). Biotinylated RNA baits (2 μg) were incubated with pre-cleared cell lysates supplemented with 0.4 units μl^{-1} RNasin (Promega) for 30 min at room temperature followed by 2 h incubation at 4 °C. The mixture was then added to streptavidin-conjugated agarose beads pre-blocked with BSA (1%) and tRNA (50 μg ml^{-1}) for 2 h at 4 °C. RNA–protein complexes were pulled-down and washed extensively. Samples were separated on 10% (w/v) polyacrylamide Bis-Tris gels (Invitrogen) and stained with Imperial protein stain (Thermo scientific). Proteins in gel slices were digested with trypsin and identified using LC-MS/MS by the Weizmann Institute's Mass Spectrometry Unit. Three independent biological replicates were performed. For western blot analyses samples were separated on 10% (w/v) polyacrylamide Bis-Tris gels (Invitrogen) and transferred onto nitrocellulose membrane using iBlot gel transfer system (Invitrogen) set to P3 for 8 min with iBlot gel transfer stacks (Invitrogen). Membranes were blocked in 5% BSA, 0.05% Tween-20 in PBS for 1 h, and then incubated overnight at 4 °C with anti-YTHDF2 or anti-ELAVL1 polyclonal antibody (Abcam) diluted 1:500 in 5% milk; anti-METTL3 (Abnova) diluted 1:2,000; anti-actin (Epitomics) diluted 1:5,000. Proteins were visualized using the SuperSignal West Pico Luminol/Enhancer solution (Thermo scientific).

Alignment of reads. Transcript-based analysis. Reference transcriptomes were prepared based on the University of California, Santa Cruz (UCSC) KnownCanonical tables in human (hg18) and mouse (mm9) consisting of a single representative transcript for each gene[49], and comprising 19,965 coding and 5,950 non-coding genes for human, and 20,437 and 6,739, respectively, for mouse. Reads were first aligned against the relevant reference using Novoalign 2.07 (Novocraft Technologies SdnBhd; http://www.novocraft.com), discarding all reads that were mapped to multiple genomic regions. The remaining reads were realigned against the relevant transcriptome and reads that aligned less well to the transcriptome than to the genome were also discarded. A strategy of iteratively trimming two nucleotides from the end of the read and then realigning it was used when a read did not align at its original length (-s 2 parameter). Trimming of reads to a length shorter than 25 nucleotides (-l 25 parameter) was not allowed. Additional parameters used for the indexing step were '-k 14 -s 2', and for the alignment was '-t 75'.

Genome-based analysis. Reads were mapped to the human genome using Bowtie[50] with '-best -strata' parameters. Non-unique reads mapping to more than ten locations were discarded from downstream analysis.

Calculation of coverage. Reads were artificially extended to a length of 100 nucleotides in the 5′-to-3′ direction, to account for the average length of RNA fragments, thus ensuring coverage of methylation sites residing outside the sequenced 36-nucleotide stretch. The extended reads were used to generate gene coverage maps summing the number of reads overlapping each nucleotide in every gene. This approach relies on the narrow fragment size range (90–150 nucleotides, measured by Agilent Bioanalyzer). Analyses were limited to human genes with at least 40 reads per 1,000 nucleotides, to avoid misinterpretations due to insufficient coverage. Expression levels were evaluated by calculating the number of reads in each gene, and dividing the range into five bins.

Detection of m^{6}A sites. Search for enriched peaks in the m^{6}A immunoprecipitation sample compared to the input control was performed by scanning each gene using sliding windows of 100 nucleotides with 50 nucleotides overlap. The mean coverage for each window was calculated for the immunoprecipitated and control samples (MeanWinIP and MeanWinControl, respectively). Gene median coverages for immunoprecipitation and input control were determined (MedianGeneIP and MedianGeneControl, respectively) to robustly estimate background levels. Every window was assigned the following unit-less metric, representing the enrichment fold change of immunoprecipitated over input control samples after normalization by background.

$$\text{winscore} = \log_2\left(\frac{\text{MeanWinIP}/\text{MedianGeneIP}}{\text{MeanWinControl}/\text{MedianGeneControl}}\right)$$

A robust estimation signal was ensured by setting a limit of mean window coverage >20 for the immunoprecipitated sample. The FDR of this method was estimated by comparing the number of enriched windows exceeding a set threshold using this approach to the number that is obtained if the experiments were computationally reversed with input being treated as immunoprecipitate and vice versa. For a threshold of 2, corresponding to a fourfold increase with respect to control, the FDR was <7%; thus, the winscore threshold was set to 2. All overlapping enriched windows were merged and positions with maximal immunoprecipitate-positive and immunoprecipitate-negative coverage were identified. If the distance between these two maxima was <100 nucleotides and the ratio between the amplitude of these two maxima was less than twofold, a putative m^{6}A site was defined in the centre of these two maxima, and a score was allocated to this peak, peakscore, corresponding to the maximal winscore from among all the overlapping windows. When these conditions were not met, windows were classified as non-resolved peaks and excluded from the analysis, to avoid PCR/sequencing artefacts. The negative control peaks were detected as described above: All windows with winscore <0 were selected. Control windows were

selected from the same genes as the detected m^{6}A peaks (limited to a mean window coverage >20). All overlapping control windows matching these criteria were merged and a negative control peak was defined in their centre.

For genome-based analysis, the peak-caller MACS[51] was used for peak detection; the 'effective genome size' parameter was adjusted to the calculated transcriptome size (1.35×10^8). Peaks were considered if their MACS-assigned fold change was >4 and individual FDR value <5%.

Assignment of m^{6}A peaks into exons representing different splicing events. m^{6}A peak summits were intersected with a data set of genomic exon coordinates belonging to ten splicing categories. A list of splicing event exons of all coding Ensembl genes (version 63 hg19) was downloaded using the BioMart tool[52]. After lifting over peak summits to hg19 coordinates, they were intersected with the downloaded splicing event list.

Identification and clustering of enriched motifs. In-house method for detecting motifs using all identified peaks. Motifs enriched within m^{6}A peaks compared with control peaks were identified by counting the occurrence of 4–6-nucleotide *k*-mers in the immunoprecipitate and its corresponding control group. The total number of *k*-mers of each length within every group was counted and the ratio between their prevalence was used to calculate the fold change between the two groups. Exact Fisher test was used to evaluate the differences in the prevalence of each *k*-mer between the groups. Analysis was limited to motifs enriched more than twofold and with an associated Bonferonni-corrected *P* value < 0.05. Motifs were clustered together, using the previously described approach[53]. To correct for the underlying base composition we repeated these analyses using a second set of control sequences based on randomly permuting the sequences of the m^{6}A sites, validating the significance of the found motif.

MEME search. MACS-identified peaks with FDR ≤ 5% were sorted according to their fold change. The top 1,000 peaks falling within known genes were chosen for *de novo* motif analysis. 101-nucleotide-long sequences derived from the sense strand and centred around the peak summit were used as input for MEME[54].

Conservation analysis. Human–mouse orthologous genes were identified based on the mmBlastTab table, which is linked to the UCSC KnownCanonical table. Orthologous sequences were then aligned using the Needleman–Wunsch global alignment algorithm implemented in 'needle'. Mouse and human m^{6}A peaks were identified between their start and end coordinates (see above) and projected onto the alignment to enable the identification of conserved m^{6}A peaks. The distance of the intervals in human was limited to be >50% the length in mouse, to ensure minimal sequence identity; sites not matching this criterion were discarded. Analyses were limited to human genes with at least 40 reads per 1,000 nucleotides, to avoid misinterpretations due to insufficient coverage.

m^{6}A stability across experiments. To explore the stability of m^{6}A peaks across different experiments, the presence of each peak identified in a particular (reference) experiment was confirmed in the other (target) experiments. Conserved peaks were defined as peaks existing in both experiments between matching coordinates. Only peaks in genes with a mean expression exceeding 40 reads per 1,000 nucleotides in the target experiment were considered, to guarantee sufficient coverage. Borderline cases passing the threshold in one experiment but not in the other were minimized by limiting the peaks in the reference experiment to a peakscore of >3. For the analysis of differential peaks between two experiments, we demanded that the peakscore be <0 in the target experiment lacking the site, to ensure the presence of a peak in only one condition.

Secondary structure. RNAfold[55] was used to assess secondary structures within 100-nucleotide sequences centred around the detected m^{6}A and control peaks with default parameters to calculate the minimum free energy (MFE) of the folded sequence. Each sequence was also shuffled 50 times and MFE scores were calculated for those sequences as well. Sequences were then assigned a *Z* score, indicative of the extent to which a sequence is more stably folded compared to the shuffled controls.

Multi-layer differential expression analysis. For all differential expressed layers, features were considered as differentially expressed only when fold change ≥ 2 and FDR ≤ 5% using Benjamini–Hochberg multiple testing adjustment.

Differential expressed genes. RNA-seq reads were mapped to the human genome (build hg18) using TopHat 1.2.0 software[56]. The number of reads mapped to each of the Ensembl genes (release 54) was counted using the HTSeq python package[57], with the 'union' overlap resolution mode, and -stranded = no. The R package DESeq v.1.5.24[58] within the Bioconductor framework was used for differential expression analysis.

Differentially expressed exons. The analysis was done using the R package DEXSeq[59] within the Bioconductor framework. The package provides a method to systematically detect differential exon usage, that is, whether the proportion of read from a given exon among all the reads that fall onto a gene is significantly changed. The analysis was done in the following way. First, non-overlapping exonic regions were defined using the 'dexseq_prepare_annotation.py' script provided as part of the DEXseq package. Next the number of reads falling in each of the defined exonic regions was counted using the DEXseq script 'dexseq_count.py' with parameters -a = 0 to include multi-reads mapped to different locations of the genome, and -stranded = no.

Differentially expressed introns. Non-overlapping intronic regions that also do not overlap any exon on either strand were defined using an in-house script. The number of reads falling in each of the defined regions was counted using a modified version of the 'dexseq_count.py' DEXseq script, and differential expression analysis was done using the DESeq Bioconductor package.

Differentially expressed transcripts. An Ensembl .gtf file of all human genes (hg18 release 54) was re-processed using Cuffcompare v.1.0.3 to add the missing tss_id and p_id attributes according to the user guide. The resulting .gtf annotation file created by Cuffcompare was used as input to Cuffdiff v.1.0.3 tool together with the fragment alignment files. Both Cuffcompare and Cuffdiff are part of the Cufflinks package[60].

Intersection between data sets. Intersection operations between genomic locations of differentially expressed features and the identified peaks were done using BEDTools[61] and PeakAnalyzer software[62]. Hypergeometric test was used for calculating enrichment *P* values, unless stated otherwise.

Statistical analysis and graphics. All statistical analyses (unless stated otherwise) were performed using the R package for Statistical Computing. Most of the presented figures were produced using the ggplot2 package[63]. Sequence logos were prepared using the SeqLogo package[64].

49. Hsu, F. *et al.* The UCSC Known Genes. *Bioinformatics* **22,** 1036–1046 (2006).
50. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10,** R25 (2009).
51. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9,** R137 (2008).
52. Kinsella, R. J. *et al.* Ensembl BioMarts: a hub for data retrieval across taxonomic space. *Database* **2011,** bar030 (2011).
53. Llorian, M. *et al.* Position-dependent alternative splicing activity revealed by global profiling of alternative splicing events regulated by PTB. *Nature Struct. Mol. Biol.* **17,** 1114–1123 (2010).
54. Bailey, T. L. & Elkan, C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc. Int. Conf. ISMB* **2,** 28–36 (1994).
55. Zuker, M. & Stiegler, P. Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Res.* **9,** 133–148 (1981).
56. Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25,** 1105–1111 (2009).
57. Anders, S. HTSeq: analysing high-throughput sequencing data with Python http://www-huber.embl.de/users/anders/HTSeq/doc/overview.html (2010).
58. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11,** R106 (2010).
59. Anders, S., Reyes, A. & Huber, W. Detecting differential usage of exons from RNA-Seq data. Available from *Nature Precedings* http://hdl.handle.net/10101/npre.2012.6837.1 (2012).
60. Roberts, A., Pimentel, H., Trapnell, C. & Pachter, L. Identification of novel transcripts in annotated genomes using RNA-Seq. *Bioinformatics* **27,** 2325–2329 (2011).
61. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26,** 841–842 (2010).
62. Salmon-Divon, M., Dvinge, H., Tammoja, K. & Bertone, P. PeakAnalyzer: genome-wide annotation of chromatin binding and modification loci. *BMC Bioinformatics* **11,** 415 (2010).
63. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* (Springer, 2009).
64. Bembom, O. *seqLogo: Sequence Logos for DNA Sequence Alignments* (Division of Biostatistics, University of California, Berkeley, 2011).