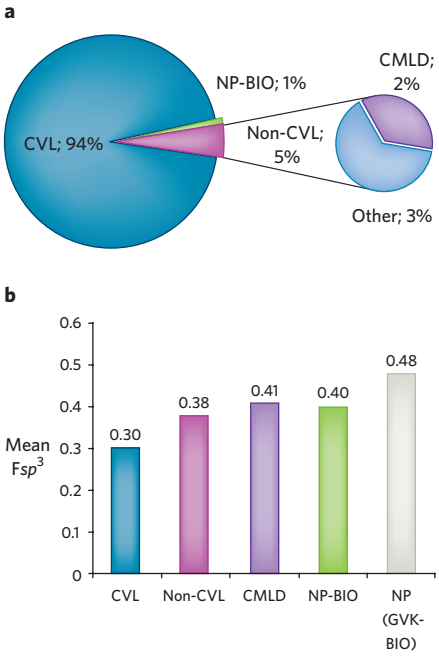


Figure 2 | Analysis of structural complexity of MLSMR collection. (**a**) Composition of 2010 MLSMR. CVL, MLSMR compounds annotated as commercial including targeted libraries (317,478 compounds); NP-BIO, MLSMR compounds annotated as natural products (NP) and bioactives (3,938 compounds); non-CVL, MLSMR compounds annotated as noncommercial (16,802 compounds); CMLD, MLSMR compounds submitted by CMLD centers (5,979 compounds). (**b**) Mean Fsp^3 values for the 2010 MLSMR (337,890 compounds) and natural products retrieved from the GVK-BIO database (NP (GVK-BIO); 2,265 compounds), where Fsp^3 = number of sp^3 hybridized carbons over total carbon count. Mean Fsp^3 values were determined according to the method of Lovering *et al.*[14]. Natural products and bioactives from the MLSMR were combined in this analysis, because many natural products are included in the bioactives category.

Even when small molecule 'hits' are identified from less trodden chemical space by academia, conducting medicinal chemistry research on these hits remains a substantial challenge. One of the contributing factors is the lag time between the synthesis and identification of the hit. In many cases, by the time the chemist who synthesized the small molecule becomes aware of the hit, he or she has moved on to another project or even advanced to the next stage of his or her professional career. In addition, many compounds from academic labs were not prepared as part of a library project, meaning ready access to a series of analogs is not available. Even in the case of compounds created for this purpose, most of the grants that enable library synthesis do not fund follow-up studies and hence resource allocation for the optimization phase becomes a challenge. Looking toward the future, funding mechanisms enabling the development of high-quality probes or drug leads resulting from neglected chemical space would reduce these hurdles. For example, collaborative research between specialized centers such as CMLDs that produce complex small molecules and MLPCNs or MLSCNs that offer high-throughput screening and follow-up secondary screening could be a logical next step.

Despite their popularity, CVLs are underperforming for many classes of biological targets. Synthetic chemists, biologists and screening centers should share information and foster collaborations toward the goal of expanding the druggable chemical space. The next generation of drug discovery relies on it. New policies or incentives by funding agencies could promote broader participation by synthetic chemists, and new funding mechanisms are needed to promote the medicinal chemistry of underexplored chemical space. Because chemical biologists focus on the interface of synthetic chemistry and cell biology, they are well equipped to lead the way in addressing these important challenges. ■

Sivaraman Dandapani and Lisa A. Marcaurelle are members of the Chemical Biology Platform at the Broad Institute of Massachusetts Institute of Technology and Harvard University, Cambridge, Massachusetts, USA.
e-mail: sivaram@broadinstitute.org or lisa@broadinstitute.org

References

1. Hüser, J., Mannhold, R., Kubinyi, H. & Folkers, G. (eds.). *High-Throughput Screening in Drug Discovery* (Wiley, 2006).
2. Payne, D.J., Gwynn, M.N., Holmes, D.J. & Pompilano, D.L. *Nat. Rev. Drug Discov.* **6**, 29–40 (2007).
3. Tepper, R.I. & Roubenoff, R. in *Genomics and Personalized Medicine* (Willard, H.F. & Ginsburg, G.S., eds.) 335–342 (Elsevier, 2009).
4. Drewry, D.H. & Macarron, R. *Curr. Opin. Chem. Biol.* **14**, 289–298 (2010).
5. Danishefsky, S. *Nat. Prod. Rep.* **27**, 1114–1116 (2010).
6. Tan, D.S. *Nat. Chem. Biol.* **1**, 74–84 (2005).
7. Dandapani, S. & Marcaurelle, L.A. *Curr. Opin. Chem. Biol.* **14**, 362–370 (2010).
8. Stanton, B.Z. *et al. Nat. Chem. Biol.* **5**, 154–156 (2009).
9. Marcaurelle, L.A., Johannes, C., Yohannes, D., Tillotson, B.P. & Mann, D. *Bioorg. Med. Chem. Lett.* **19**, 2500–2503 (2009).
10. Rottmann, M. *et al. Science* **329**, 1175–1180 (2010).
11. Workman, P. & Collins, I. *Chem. Biol.* **17**, 561–577 (2010).
12. Blake, J.F. *Curr. Opin. Chem. Biol.* **8**, 407–411 (2004).
13. Shelat, A.A. & Guy, R.K. *Curr. Opin. Chem. Biol.* **11**, 244–251 (2007).
14. Lovering, F., Bikker, J. & Humblet, C. *J. Med. Chem.* **52**, 6752–6756 (2009).
15. Clemons, P.A. *et al. Proc. Natl. Acad. Sci. USA* published online, doi:10.1073/pnas.1012741107 (18 October 2010).

Competing financial interests

The authors declare no competing financial interests.

RNA epigenetics?

Chuan He

Post-transcriptional RNA modifications can be dynamic and might have functions beyond fine-tuning the structure and function of RNA. Understanding these RNA modification pathways and their functions may allow researchers to identify new layers of gene regulation at the RNA level.

For a long time, DNA took first place in biological research on nucleic acids. However, the discoveries of catalytic RNAs and functional noncoding RNAs over the past two or three decades have completely changed our views on RNA. RNA research has become one of the most dynamic and fast-growing fields in science. Cellular RNAs (ribosomal RNAs (rRNAs), transfer RNAs (tRNAs), messenger RNAs (mRNAs), small nuclear RNAs (snRNAs) and others) contain more than a hundred structurally distinct post-transcriptional modifications at thousands of sites (http://rna-mdb.cas.albany.edu/RNAmods/). Many of these modifications have been known for decades. However, despite this large chemical diversity and the rapid advance

of RNA research, the enzymes that post-transcriptionally modify RNA and the roles of covalent changes of RNA have been less investigated. A common view of RNA modification is that it has an adaptive role that can fine-tune the structures and functions of mature RNAs to influence gene expression[1], but recent studies have also hinted that RNA modification may have other functions. Some post-transcriptional RNA modifications can be dynamic and might have regulatory roles analogous to those of post-translational protein modifications. Understanding the scope and mechanisms of these dynamic RNA modifications, which could be termed 'RNA epigenetics', represents a new frontier in research for chemical biologists.

The most common and simple RNA modifications involve the methylation of bases and 2′-hydroxyls of RNA nucleotides (**Fig. 1a**), but the functions and biological consequences of most methylated RNAs are unknown. For example, in all higher eukaryotes each RNA molecule has three to five 6-methyladenine (6-meA) bases. Interestingly, knockdown of the methyltransferase that is responsible for 6-meA formation in HeLa cells using small inhibitory RNA (siRNA) leads to apoptosis[1], but the exact function and molecular actions of 6-meA in mRNA remain unclear. The roles of a few other RNA base methylations have been more clearly elucidated. The 1-methyladenine (1-meA) modification at A58 is conserved in most eukaryotic tRNAs and is essential for viability in yeast[2]. However, there is now evidence that this modification is not as static as has been assumed: around 25% of human tRNA species are hypomodified at this conserved nucleotide in several human cell lines[3]. Many of the methyltransferase enzymes that install these methyl groups have been identified, but there continue to be surprises. For instance, DNMT2, a well-known DNA methyltransferase, has recently been shown to catalyze the methylation of C38 in the anticodon loop of tRNA[4], which might modulate the folding and stability of tRNAs and regulate protein synthesis under stress conditions[5]. Certain RNA methyltransferases can catalyze the methylation of adenine in ribosomal RNA to yield 2,8-dimethyladenine (2,8-dimeA)[6]; this activity has been discovered in a hospital isolate of *Staphylococcus aureus*, which leads to resistance of this prevalent pathogen to several commonly used antibiotics[7].

This view of dynamic RNA modification is supported by the discovery of several classes of enzymes that seem to operate on modified and natural RNA bases. A group of mononuclear iron-containing dioxygenases—including FTO, ABH1 and ABH3—have been shown to catalyze the oxidative demethylation of methylated bases that are probably located in single-stranded RNA[8–10] (**Fig. 1b**). ABH8 is a related enzyme that catalyzes the hydroxylation of a modified uridine (mcm^5U) in the wobble position of the anticodon stem loop of a specific tRNA[11] (**Fig. 1b**). These modifying enzymes have been linked to important biological pathways, particularly in cancer cells[12–14]. However, FTO, which was identified in 2007 from genome-wide association studies as an obesity susceptibility factor[15–17], offers a particularly intriguing example. FTO is a DNA and RNA demethylase that affects body mass index and development through an unknown mechanism[18]. Although the RNA-processing function of FTO has been suggested to affect gene expression, its substrates and mechanisms of action await further elucidation. Similarly, JMJD6—another iron-containing hydroxylase, which was originally characterized as a histone demethylase and lysine hydroxylase[19]—also binds efficiently to single-stranded RNA and might have important but unknown roles in processing or modifying RNA[20].

Together, these recent findings support a more dynamic view of RNA modification states, which substantially increases the complexity and diversity of RNA species in cells and may add another layer to the regulation of gene expression at the RNA level. This opens up a key frontier for chemical exploration of the functions and mechanisms of these pathways.

How will chemical biologists advance this area of science? First, as mutations in FTO and related dioxygenases cause intriguing phenotypes, it is important to identify the cellular substrates of these enzymes and to understand their mechanisms (**Scheme 1**). Chemical biologists can make key contributions by inventing new tools. CLIP-Seq is a good example, in which potential RNA substrates are identified through cross-linking to the target RNA-binding protein, immunoprecipitation to capture the mixture and high-throughput sequencing

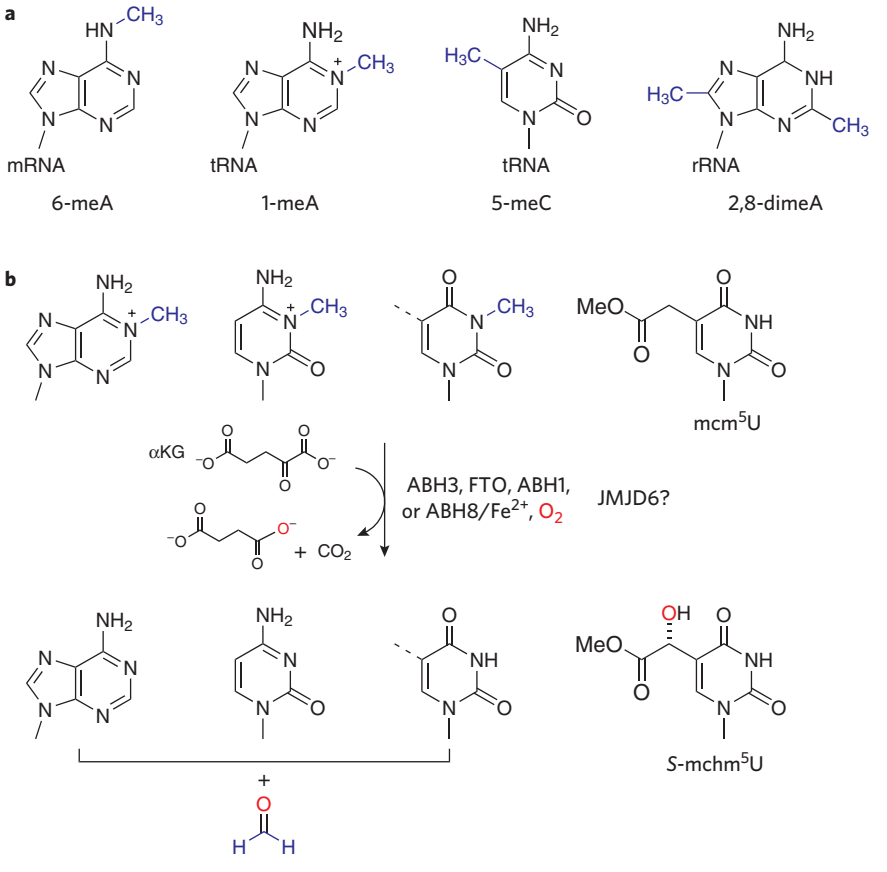
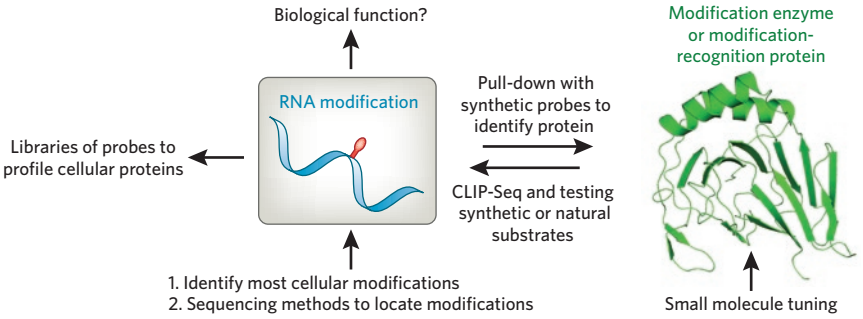


Figure 1 | Examples of RNA modification and demodification that may impact biological regulation. (**a**) Selected examples of RNA base methylation. (**b**) A group of dioxygenases that use iron, α-ketoglutarate and dioxygen to perform oxidation of modified RNA bases for demethylation or hypermodification.

for RNA identification. In addition, chemical synthesis of potential substrates to validate enzymatic targets is essential, as shown in the study of ABH8 (ref. 11). Once the potential RNA substrate has been revealed, its connection to biological regulation needs to be established. Again, using ABH8 as a specific example, the hydroxylation of mcm^5U in the wobble position of the anticodon stem loop creates a new hydrogen bond in the modified uracil (**Fig. 1b**), which might affect wobble recognition and modulate translation through codon usage. Besides translation, RNA splicing and RNA transport could be affected through the enzymatic addition or reversal of these modifications, creating new layers of regulation. For instance, FTO or related dioxygenases might have cellular substrates that are linked to RNA splicing or transport, and the identification of their substrates, through CLIP-seq or newly invented techniques, will be key to the elucidation of their functional roles *in vivo*.

Second, chemical biologists can contribute by clarifying the RNA modification reactions. Before we can elucidate the exact functions and molecular mechanisms of known modifications such as 6-meA in mRNA, we will need to develop chemical, biochemical or single-molecule-based sequencing methods so that we can identify the locations of these modifications. Synthetic RNA probes can be prepared for noncovalent or covalent (for example, using photoactive cross-linkers) pull-down experiments, which can be used to identify proteins that might specifically recognize and process the modifications (**Scheme 1**). Eventually, systematic studies of RNA modification by cellular machineries and a system-wide understanding of the dynamics of these modifications will be required if we are to fully appreciate the functional roles of modified RNA. With a comprehensive knowledge of the diverse RNA modifications that are found in mammalian cells, chemical biologists might be able to prepare a range of synthetic probes carrying these modifications that can be used to profile cellular proteins that recognize specially modified RNAs (**Scheme 1**). This will provide essential tools for understanding the components and global mechanisms of these pathways.



Scheme 1 | Chemical biology of RNA epigenetics. Reversible RNA modifications, their roles in biology and proteins that recognize and/or process these modifications will need to be identified and characterized. Chemical biology can have a key role in this research.

Finally, by analogy to the extensive work on histone modification and demodification systems, chemical biologists will lead the effort to identify small-molecule inhibitors for RNA modification or demodification enzymes involved in human diseases, which will serve as powerful probes for these pathways and potential leads for new therapies (**Scheme 1**). These efforts will rely heavily on synthetic chemistry and on the ability of chemical biologists to apply these compounds in real biological settings. In addition, if robust cell-based assays can be developed to perturb certain RNA modifications that influence biological regulation, the cellular components that are involved in these processes might be identified using chemical genetic approaches.

Given the rich layers of epigenetic regulation that result from targeted modifications of DNA and proteins, reversible RNA modification might represent another realm for biological regulation in the form of 'RNA epigenetics'. Understanding the underlying pathways and mechanisms of this regulation might change the way we think about gene expression and development. ■

Chuan He is in the Department of Chemistry and Institute for Biophysical Dynamics, University of Chicago, Chicago, Illinois, USA.
e-mail: chuanhe@uchicago.edu

References

1. Grosjean, H. (Ed.) *Topics in Current Genetics* Vol. 12 (Springer, Berlin, 2005).
2. Anderson, J. *et al. Genes Dev.* **12**, 3650–3662 (1998).
3. Saikia, M. *et al. RNA* **16**, 1317–1327 (2010).
4. Goll, M.G. *et al. Science* **311**, 395–398 (2006).
5. Schaefer, M. & Lyko, F. *Chromosoma* **119**, 35–40 (2010).
6. Yan, F. *et al. J. Am. Chem. Soc.* **132**, 3953–3964 (2010).
7. Toh, S.M. *et al. Mol. Microbiol.* **64**, 1506–1514 (2007).
8. Yi, C., Yang, C.G. & He, C. *Acc. Chem. Res.* **42**, 519–529 (2009).
9. Trewick, S.C. *et al. Nature* **419**, 174–178 (2002).
10. Falnes, P.Ø., Johansen, R.F. & Seeberg, E. *Nature* **419**, 178–182 (2002).
11. Fu, Y. *et al. Angew. Chem. Int. Ed.* published online, doi:10.1002/anie.201001242 (25 June 2010).
12. Shimada, K. *et al. Cancer Res.* **69**, 3157–3164 (2009).
13. Aas, P.A. *et al. Nature* **421**, 859–863 (2003).
14. Shimada, K. *et al. Cancer Sci.* **99**, 39–45 (2008).
15. Frayling, T.M. *et al. Science* **316**, 889–894 (2007).
16. Dina, C. *et al. Nat. Genet.* **39**, 724–726 (2007).
17. Scott, L.J. *et al. Science* **316**, 1341–1345 (2007).
18. Gerken, T. *et al. Science* **318**, 1469–1472 (2007).
19. Webby, C.J. *et al. Science* **325**, 90–93 (2009).
20. Hong, X. *et al. Proc. Natl. Acad. Sci. USA* **107**, 14568–14572 (2010).

Competing financial interests

The author declares no competing financial interests.