doi:10.1038/nature24041

Supplementary Note 1. Identifying new RNA editing sites from human and mouse RNA-seq data

We and others have recently identified a large number of RNA editing sites in human and mouse[2-7,13]. Despite the great success, it is unlikely that the discovery of editing sites has been saturated, especially in the mouse where few studies have been reported. To identify additional editing sites, we constructed RNA-seq libraries from multiple human adult and fetal tissues as well as multiple inbred mouse adult and embryonic tissues (Supplementary File 9). We also constructed exome sequencing libraries for the same human tissues and used publicly available genome sequences of inbred mice[47,48,60] to differentiate SNPs from genuine editing sites. In addition, we took 318 the most deeply sequenced GTEx samples spanning from age 20 to 70 and covering both genders evenly from 53 body sites, together with whole-genome sequencing (WGS) of the donors to identify RNA editing sites. By applying our in-house computational pipeline[6,7] to the RNA-seq and exome-seq/WGS libraries as well as to publicly available RNA-seq libraries for mouse tissues[3], we identified 185,435 new A-to-I editing sites in our human samples, 193,523 new sites in GTEx samples and 31,347 new sites in mouse. A-to-G mismatches, indicative of A-to-I editing, comprised over 70% of all mismatches called between DNA and RNA, indicating that the estimated false discovery rate (FDR) is <3%. After including the sites that are currently curated in the RNA editing database RADAR[5], the total number of A-to-I editing sites in human and mouse identified to date are 2,802,751 and 38,504 respectively. We validated 249 of 314 randomly selected human sites (79.3%) in 80 genes and 135 of 149 mouse sites (90.6%) in 35 genes by Sanger sequencing (http://lilab.stanford.edu/atlas). The true positive rate is likely to be higher, since most sites that could not be validated are edited at low levels that would be missed due to the poor sensitivity of Sanger sequencing. In addition, some non-validated sites may be edited exclusively in tissues not used for the validation.

Supplementary Note 2. Identifying new RNA editing sites from human and mouse mmPCR-seq data

We recently developed a method that combines microfluidic multiplex PCR and deep sequencing (mmPCR-seq) to accurately and efficiently measure RNA editing levels at hundreds of editing sites located in 240 human loci[21]. Here, we expanded mmPCR-seq to amplify 672 human and 557 mouse exonic loci, each of which contains one or more editing sites (see

Supplementary File 3 for primer sequences). All known non-repetitive sites in protein coding regions were included. Editing level measurements by mmPCR-seq were highly reproducible and consistent with measurements by RNA-seq, agreed well with Sanger sequencing, and were unaffected by the complexity of the PCR or the choice of reverse transcriptases (data not shown). Hence, mmPCR-seq can accurately quantify the editing levels of most genes, except some that are expressed at very low levels.

Because A-to-I editing sites often occur in clusters, we wondered whether we could identify additional RNA editing sites from the PCR amplicons. By requiring >80% of mismatches called between DNA and RNA to be A-to-G for every sample, we identified 7,310 new human and 9,921 new mouse editing sites, with an estimated FDR of ~2%. The vast majority of these new sites are edited at low levels, compared to the sites in curated genome-wide datasets. Nevertheless, we validated 42 new sites in 13 human genes and 49 new sites in 14 mouse genes by Sanger sequencing (http://lilab.stanford.edu/atlas). By combining all the queried sites and the novel sites, we could accurately quantify the editing levels for 12,871 human sites and 11,103 mouse sites using mmPCR-seq.

Supplementary Note 3. Co-edited and exclusively edited events in human tissues

Sites that display tissue-specific editing patterns may perform important functions pertinent to those tissues. We found that 2,094 sites exhibited highly variable editing in the GTEx datasets (coefficient of variation > 0.8). Co-editing network analysis of these variable sites revealed that the samples segregated into three major tissue groups and that the editing sites could be divided into seven clusters (Extended Data Fig. 1d). Gene Ontology (GO) analysis indicated that the brain-specific cluster was enriched for sites located in genes related to glycoprotein metabolism, which suggests a role for RNA editing in neurodegenerative diseases where aberrant glycosylation is a contributory factor. Additionally, GO analysis showed that sites that were edited exclusively in the blood vessels (Extended Data Fig. 1e and Supplementary File 4) were often found in genes encoding transcription factors, such as HIF1A and HIF3A, which are important regulators of vascularization and angiogenesis[61].

Supplementary Note 4. RNA editing analysis in mouse tissues

We profiled A-to-I editing in multiple mouse tissues using mmPCR-seq. Consistent with previous studies, the mouse brain showed the highest overall editing level compared to non-brain tissues (Extended Data Fig. 3a). Overall editing levels were also positively correlated with the expression of *ADAR1* and *ADAR2* ($R^2 = 0.875$) (Extended Data Fig. 3b) and that the editing levels of over 40% of mouse sites correlated well with *ADAR* expression levels. Expectedly, the editing profiles in different brain regions were highly correlated with one another but poorly correlated with those in non-brain tissues (Extended Data Fig. 3c). Despite this, numerous sites were still differentially edited in different brain regions (Extended Data Fig. 3d-e). We confirmed the observed editing patterns in additional mouse individuals of a different genetic background (Extended Data Fig. 3f-g) and noted that the few differences in editing between strains could be explained by local alterations in RNA structure due to single nucleotide polymorphisms (SNPs) (Extended Data Fig. 3h-i).

Supplementary Note 5. Identifying ADAR1 and ADAR2 targets in human cells

To delineate ADAR1 and ADAR2 targets in human cells, we (1) measured editing levels in 2fTGH parenteral cells either untreated or treated with IFN-α, which induces the expression of the ADAR1 p150 isoform[32], using mmPCR-seq (Extended Data Fig. 7a); (2) overexpressed each ADAR deaminase in HEK293T cells (Extended Data Fig. 7b-c) and examined editing using RNA-seq; and (3) analyzed various published ADAR1 knockdown RNA-seq data[2,33-36] (Extended Data Fig. 7d). Intotal, we curated 9,352 and 1,403 sites that are edited by ADAR1 and ADAR2 respectively, including 262 sites by both (Supplementary File 5).

REFERENCES

60. Zheng, M. *et al.* The role of Abcb5 alleles in susceptibility to haloperidol-induced toxicity in mice and humans. *PLoS Med.* **12,** e1001782 (2015).
61. Krock, B. L., Skuli, N. & Simon, M. C. Hypoxia-induced angiogenesis: good and evil. *Genes Cancer* **2,** 1117–1133 (2011).

Figure 4b

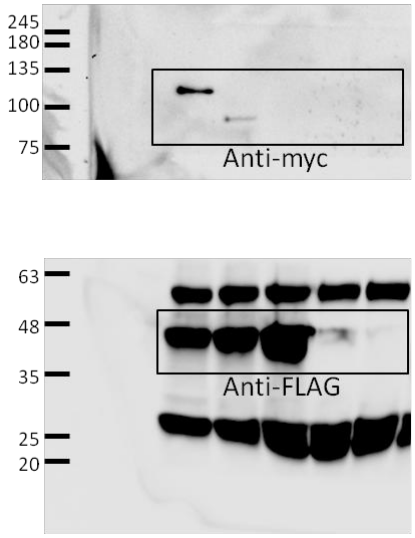


Figure 4d

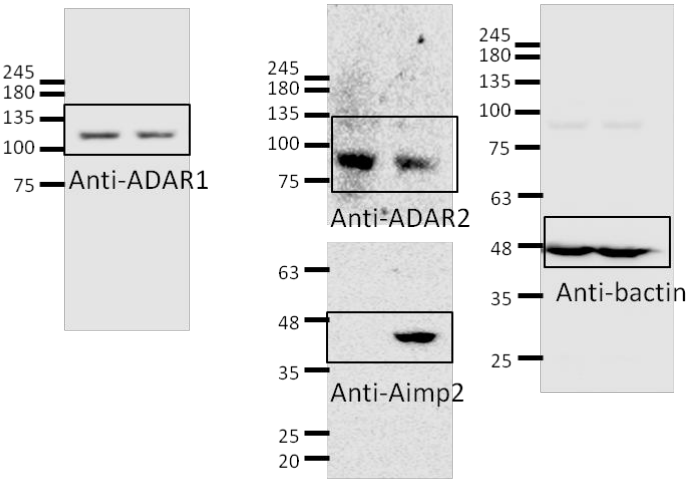
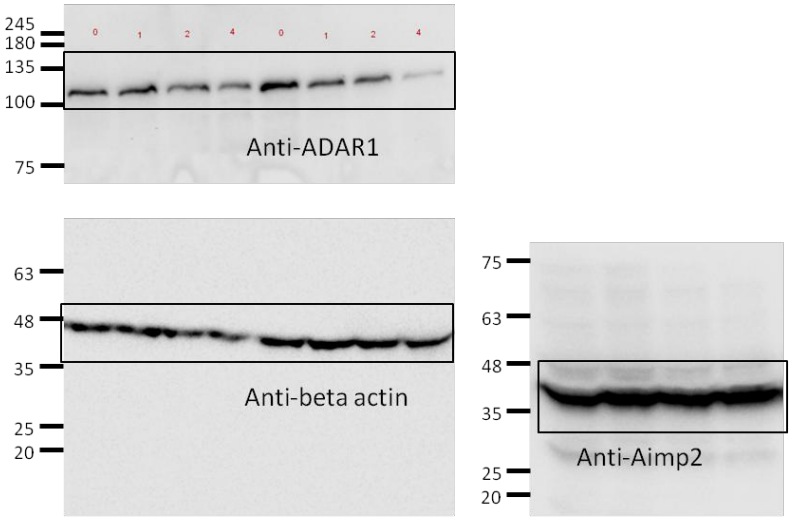
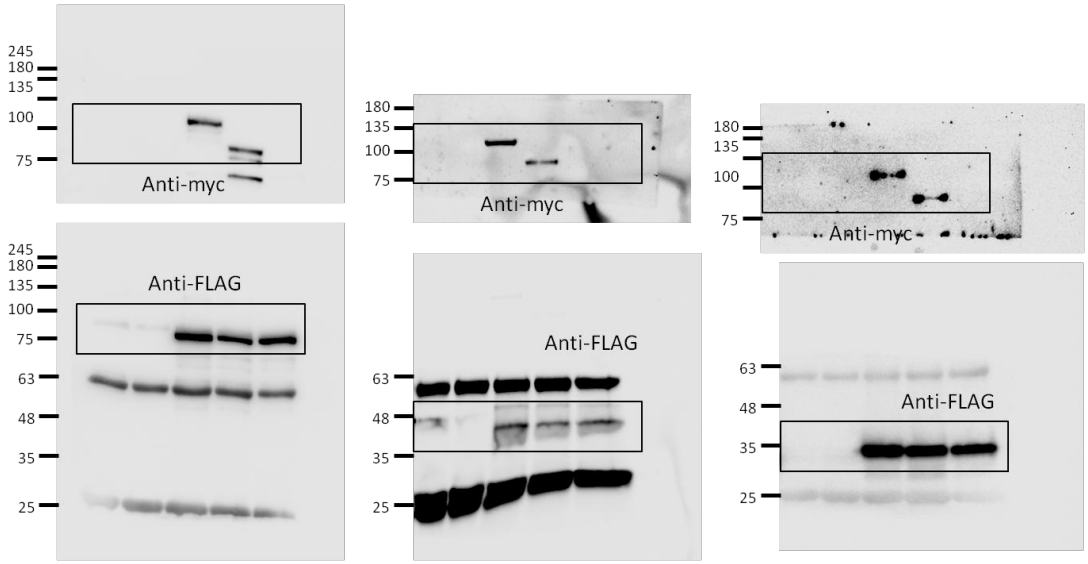


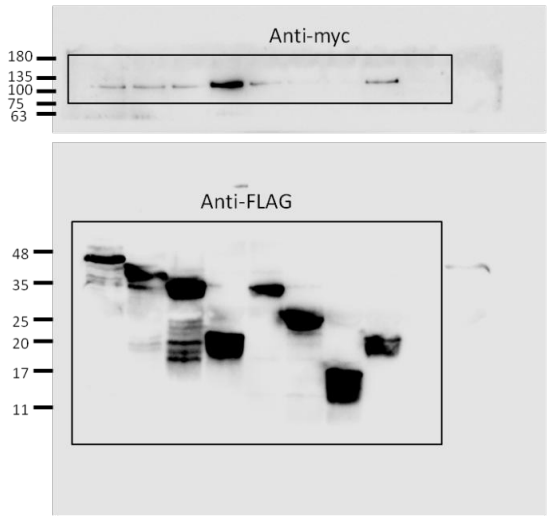
Figure 4e



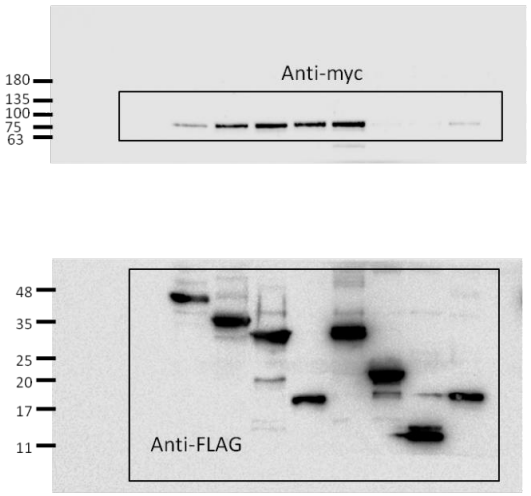
Ext Data Fig 9e



Ext Data Fig 10b



Ext Data Fig 10c



Ext Data Fig 10d