

Single-cell RNA editing identifies aggressive and favorable cellular states in chronic myelomonocytic leukemia

by Nisansala Wickramasinghe, Doan Bui, Surendra Neupane, Meghan Ferrall-Fairbanks, Michael Deininger, Eric Padron and Tongjun Gu

Received: April 3, 2026.

Accepted: August 24, 2026.

Citation: Nisansala Wickramasinghe, Doan Bui, Surendra Neupane, Meghan Ferrall-Fairbanks, Michael Deininger, Eric Padron and Tongjun Gu. Single-cell RNA editing identifies aggressive and favorable cellular states in chronic myelomonocytic leukemia.

Haematologica. 2026 Sept 3. doi: 10.3324/haematol.2026.301007 [Epub ahead of print]

Publisher's Disclaimer.

E-publishing ahead of print is increasingly important for the rapid dissemination of science.

Haematologica is, therefore, E-publishing PDF files of an early version of manuscripts that have completed a regular peer review and have been accepted for publication.

E-publishing of this PDF file has been approved by the authors.

After having E-published Ahead of Print, manuscripts will then undergo technical and English editing, typesetting, proof correction and be presented for the authors' final approval, the final version of the manuscript will then appear in a regular issue of the journal.

All legal disclaimers that apply to the journal also pertain to this production process.

Single-cell RNA editing identifies aggressive and favorable cellular states in chronic myelomonocytic leukemia

Nisansala Wickramasinghe¹, Doan Bui¹, Surendra Neupane², Meghan C. Ferrall-Fairbanks^{3,4}, Michael M Deininger^{1,5}, Eric Padron², Tongjun Gu^{1,6,7*}

¹Versiti Blood Research Institute, Milwaukee, Wisconsin

²Moffitt Cancer Center, Tampa, Florida

³J. Crayton Pruitt Family Department of Biomedical Engineering, University of Florida, Gainesville, Florida

⁴University of Florida Health Cancer Institute, University of Florida, Gainesville, Florida

⁵Department of Internal Medicine Division of Hematology and Oncology, University of Michigan Medical School, Ann Arbor, Michigan

⁶Data Science Institute, Medical College of Wisconsin, Milwaukee, Wisconsin

⁷Department of Biostatistics, University of Florida, Gainesville, Florida

*Corresponding author:

Tongjun Gu, PhD

Versiti Blood Research Institute

8727 W Watertown Plank Rd, Milwaukee, WI 53226

Email: tgu@versiti.org

Office: 414-937-3884

Acknowledgements

We thank our colleagues, Drs. Subramaniam Malarkannan, Demin Wang, and Renren Wen (Versiti Blood Research Institute), for their thoughtful discussions and suggestions. We thank Emily Diaz, PhD, for assistance with the preparation of the manuscript.

Authors' contributions

TG conceived and designed this study and performed RNA editing sites discovery. NW performed the sc RNA editing analysis. TG and NW drafted the manuscript. DB maintained and refined the RNA editing discovery pipeline and performed analyses of the two external healthy cohorts and two CMML cohorts. SN and EP provided the discovery and validation sc RNA-seq datasets. EP, MW, SN, and MF contributed critical discussions. All authors reviewed, edited and approved the final version of the manuscript.

Conflict of interests

The authors declare no competing financial interests related to this work.

Availability of data and materials

The discovery single-cell RNA-sequencing (scRNA-seq) dataset and associated clinical annotations are publicly available in the Gene Expression Omnibus (GEO) under accession GSE211033. Two additional CMML scRNA-seq validation cohorts are publicly available through the European Nucleotide Archive (ENA) under accession PRJEB37215 and GEO under accession GSE218390. Three healthy donor scRNA-seq datasets are available through GEO under accessions GSE133181, GSE218390, and GSE120221.

The code used for sc data analysis is available at <https://github.com/tjgu/CMML-scRNAediting/tree/main>.

Chronic myelomonocytic leukemia (CMML) is a clinically heterogeneous myeloid malignancy with limited therapeutic options and suboptimal risk stratification.(1, 2) Although single-cell (sc) RNA sequencing has refined CMML classification through gene expression profiling, post-transcriptional regulatory layers remain poorly defined. Adenosine-to-inosine (A-to-I) RNA editing is a pervasive RNA modification catalyzed primarily by *ADAR1* and *ADAR2*,(3) but its role in CMML at sc resolution has remained unclear(4) because accurate detection from sc RNA sequencing (scRNA-seq) is technically challenging.(5-7) Here, we show that RNA editing defines reproducible cellular states in CMML that are largely independent of gene expression-based clustering and that distinguish aggressive from clinically favorable disease states.

To enable this analysis, we developed a sc-aware framework for high-confidence RNA editing analysis that integrates stringent pseudo-bulk site discovery with sensitive sc quantification, dual alignment, cell barcode correction, per-cell duplicate removal, and multiple artifact filters. Applying this strategy to a discovery cohort of 24 CMML samples and an independent validation cohort of 13 de-identified samples, the analysis was conducted in accordance with the ethical rules and regulations of the United States.(8) We identified and validated recurrent editing-defined cellular states across cohorts. This framework provides a robust strategy for studying RNA editing at sc resolution in hematologic malignancies.

Across 56,940 cells in the discovery cohort, we identified 3,326 recurrent high-confidence A-to-I editing sites, all of which were registered in REDportal,(9) a curated RNA editing database, supporting the robustness of our site-discovery pipeline. Most editing sites were detected in only a small subset of cells (<1,000 cells), and most cells harbored only a small fraction of these sites (<50 sites). However, a subset of editing sites was detected in more than 10,000 cells, indicating that RNA editing activity is heterogeneous yet structured across the CMML cellular landscape. Annotation of these editing sites identified 961 host genes. Most editing events were located within intronic regions (~80%) and 3' untranslated regions (3'UTRs; ~12%), consistent with previous studies in other cancers.(10-12) Genes harboring intronic sites were enriched for metabolic processes, stress responses, and intracellular signaling pathways, whereas genes containing 3'UTR editing sites were specifically enriched for innate immune response and cytokine production pathways. A small number of editing sites (n=15) were also located within miRNA genes or their host transcripts. Although most genes contained one or two editing sites, 70 genes harbored at least 10. These highly edited genes were enriched for functions related to leukocyte-mediated immunity and regulation of signaling, consistent with CMML biology. Notably, genes with 3'UTR editing sites showed significantly stronger enrichment for miRNA-related terms than genes with intronic editing sites (Fisher exact test, p -value < 2.2×10^{-16}). Together, these analyses define a structured sc RNA editing landscape in CMML characterized by heterogeneous cellular distribution, distinct genomic localization patterns, and functionally coherent enrichment signatures.

We next asked whether RNA editing alone could define cellular organization. Unsupervised clustering based solely on editing profiles identified nine major editing-defined clusters (edClu0 to edClu8), with additional substructure within edClu1 (Figure 1a). To evaluate whether these RNA editing-defined clusters corresponded to biologically and clinically meaningful cellular states, we leveraged the previously established gene expression-based classification from the same CMML samples reported by Ferrall-Fairbanks *et al.*(8) That study identified a clinically aggressive granulocyte-monocyte progenitor (GMP)-like cluster, geClu2, a subset of the GMP cell type associated with poor prognosis. By transferring both cell-type annotations and gene expression cluster labels from that study onto our RNA editing-defined cells, we observed that edClu1 was predominantly composed of GMP-like cells (Figure 1b). Notably, a subregion of edClu1 showed substantial overlap with the adverse geClu2 transcriptional cluster (Figure 1c) and further subclustering identified edClu1_sub0 as the subpopulation most strongly overlapping geClu2 (Figure 1d). Marker genes defining geClu2 were significantly enriched in edClu1_sub0, and editing events within those genes were also enriched in this state (Figure 1e-1f), indicating that RNA editing independently recapitulates a clinically aggressive GMP-associated cellular program.

In contrast, edClu3 and edClu6 tracked with a more favorable disease context. Both clusters were enriched in megakaryocyte-erythroid progenitor-like (MEP) cells and shared editing programs linked to metabolism, RNA regulation, and protein modification. After hypomethylating agent treatment (HMA), edClu3 and edClu6, together with edClu2, edClu4, and edClu5, increased in normal-like samples defined previously(13) (Figure 2a-2b). In untreated samples, edClu3 and edClu6 were more abundant in CMML-0 than in CMML-1/2 and exhibited lower cellular proportions and lower mean editing levels in *TET2*-mutant disease (Figure 2c-2f). Together, these

findings suggest that these editing programs are dynamically remodeled by therapy and may reflect restoration of less aggressive hematopoietic states.

By contrast, edClu1_sub0 showed the opposite clinical pattern. In monocyte-biased samples, edClu1_sub0 demonstrated increased cellular proportion and increased mean editing level in CMML-1/2 compared with CMML-0, and in *TET2*-mutant relative to non-*TET2*-mutant disease (Figure 3a-3d). These findings further support edClu1_sub0 as an aggressive editing-defined state aligned with advanced CMML biology.

The prognostic analyses reinforced this dichotomy. In Cox proportional hazards models adjusted for age and sex and restricted to untreated patients, edClu3 showed the most consistent favorable association across models using either cluster proportion or mean editing level (p -value < 0.05). EdClu6 demonstrated a secondary but directionally concordant protective association, reaching significance only in models based on mean editing level. Additional exploratory multivariable analyses incorporating WHO subtype showed similar trends.

These findings suggest that RNA editing-defined cellular states provide complementary biological information beyond conventional clinical and molecular features used for CMML risk stratification. Although the current cohort size limits formal integration with existing prognostic models, the associations of edClu1_sub0, edClu3, and edClu6 with disease stage, treatment response, and survival indicate that RNA editing may capture additional disease heterogeneity relevant to clinical outcome.

Collectively, these analyses identify edClu3 as the strongest favorable editing-defined state and edClu6 as a secondary protective state, both associated with less advanced disease, *TET2* wild-type, improved clinical outcomes, and dynamic modulation in response to epigenetic therapy. In contrast, edClu1_sub0 is associated with more advanced CMML and adverse clinical outcomes.

To identify editing events linked to malignant hematopoiesis, we classified cells as cancer-like or healthy-like using inferCNV, with healthy reference cells from three healthy individuals, yielding 7,873 cancer-like cells and 6,325 healthy-like cells. Differential editing analysis using binomial regression identified 167 sites enriched in cancer-like cells and 109 enriched in healthy-like cells; in addition, 428 sites were unique to cancer-like cells and 70 to healthy-like cells. Genes harboring cancer-enriched editing sites were associated predominantly with immune-related and defense-response processes, whereas healthy-like-enriched sites were linked to RNA metabolic and intracellular signaling pathways. We then integrated three approaches to find high-confidence oncogenic editing events: enrichment in edClu1_sub0, enrichment in cancer-like cells, and positive association with monocyte burden (Online Supplementary Figure 1a). This yielded 92 high-confidence GMP-specific oncogenic editing sites mapping to 44 genes. These genes were enriched for immune-related pathways (Online Supplementary Figure 1b). Notably, *LAPTM5*, *CTSS*, and *CD83* harbored multiple prioritized editing sites and showed coordinated increases in both RNA editing and gene expression within edClu1_sub0 (Online Supplementary Figure 1c-1n), nominating them as candidate effectors of the aggressive CMML editing program.

Because *ADAR1* and *ADAR2* are the principal mammalian enzymes mediating A-to-I RNA editing, we next examined their relationship to global editing activity. In pseudo-bulk analyses, *ADAR1* and *ADAR2* expression were inversely correlated (Online Supplementary Figure 2a). Given that a major *ADAR1* isoform is interferon inducible,⁽³⁾ we curated a panel of 127 interferon-related and interferon-stimulated genes and incorporated the first two principal components (PCs) derived from this gene set into the correlation analysis. *ADAR1* expression was positively correlated with the first PC, whereas *ADAR2* expression was correlated with the second PC (Online Supplementary Figure 2a), suggesting distinct regulatory contexts. Although sample-level mean editing ratios tended to correlate positively with *ADAR1* and negatively with *ADAR2*, these associations were not statistically significant (Online Supplementary Figure 2b-2c), indicating that global RNA editing activity is not fully explained by bulk ADAR expression alone.

At the sc level, edClu1_sub0 showed high *ADAR1* and relatively low *ADAR2* expression (Online Supplementary Figure 2d-2e), and this pattern was independently reproduced in the validation cohort. Correlation analyses between ADAR expression and cluster-level editing activity further showed that *ADAR1* was strongly positively associated with editing activity in edClu1_sub0, edClu3, and edClu6 (p -value < 0.0001), whereas *ADAR2* showed weak associations in edClu3 and edClu6 (p -value > 0.88) and a negative association in edClu1_sub0 (p -value $\approx$ 0.012) (Online Supplementary Figure 2f-2g). Together, these findings support a model in which

differential ADAR usage contributes to distinct CMML editing programs, with an *ADAR1*-dominant, *ADAR2*-low configuration characterizing the aggressive edClu1_sub0 state.

Finally, we assessed reproducibility in an independent validation cohort of 13 CMML samples. Across 25,110 cells, we identified 1,308 editing sites and eight validation editing-defined clusters. Despite the lower number of mono-biased patients in the validation cohort, the adverse transcriptional cluster geClu2 again mapped predominantly to a GMP-enriched editing cluster. Reference mapping and marker-site validation demonstrated clear correspondence between validation and discovery states, including valClu5 with edClu1_sub0, valClu4 with edClu3, and valClu7 with edClu5/6. The key clinical patterns were also preserved: favorable editing states increased after HMA treatment and were more abundant in less advanced disease, while aggressive-state targets such as *LAPTM5*, *CTSS*, and *CD83* showed similar coordinated editing-expression patterns (Online Supplementary Figure 3a-3j). In addition, analysis of two healthy hematopoietic cohorts and two CMML cohorts(14, 15) with low proportions of monocytes and GMPs demonstrated persistent detection of edClu3, whereas edClu1_sub0 was absent or detected at very low abundance, further supporting the reproducibility of these editing-defined cellular states.

Limitations of this study include the modest sample size in some clinical and molecular subgroups, which limited mutation-stratified analyses of recurrent CMML genetic alterations, as well as the need for orthogonal validation and functional studies of key editing events. Nonetheless, our findings show that RNA editing defines reproducible and clinically relevant cellular states in CMML that are largely independent of gene expression-based clustering. A GMP-like *ADAR1*-high/*ADAR2*-low editing state aligns with aggressive disease biology, whereas edClu3 and edClu6 mark more favorable states associated with treatment response and improved survival.

References

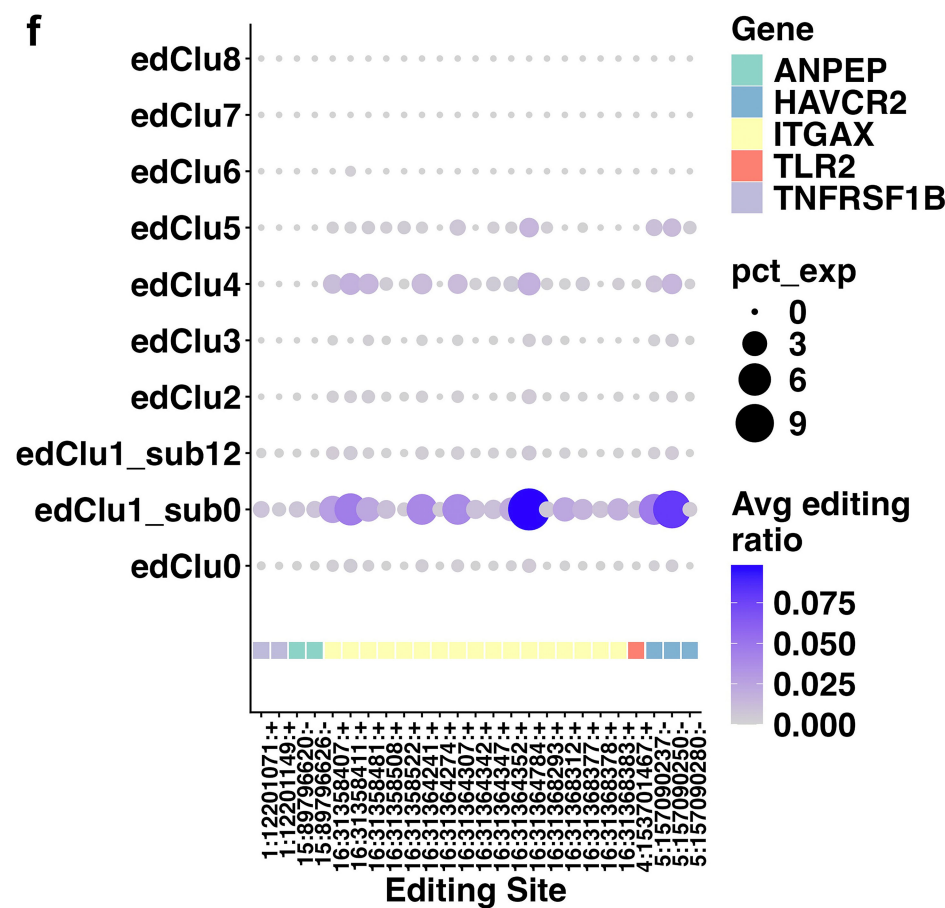
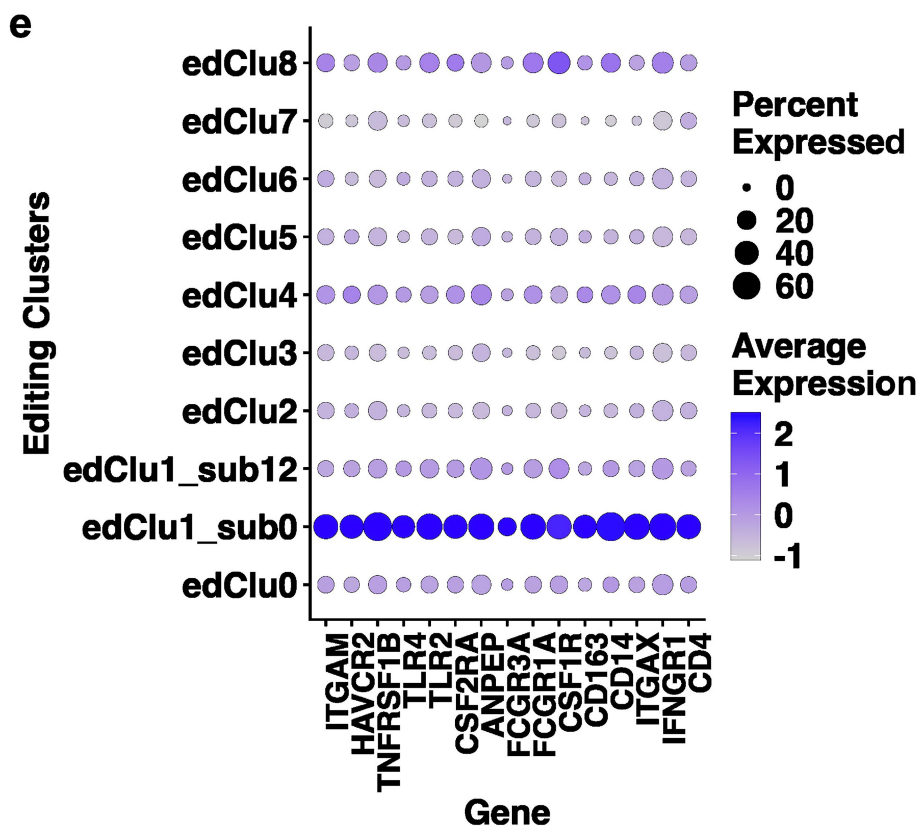
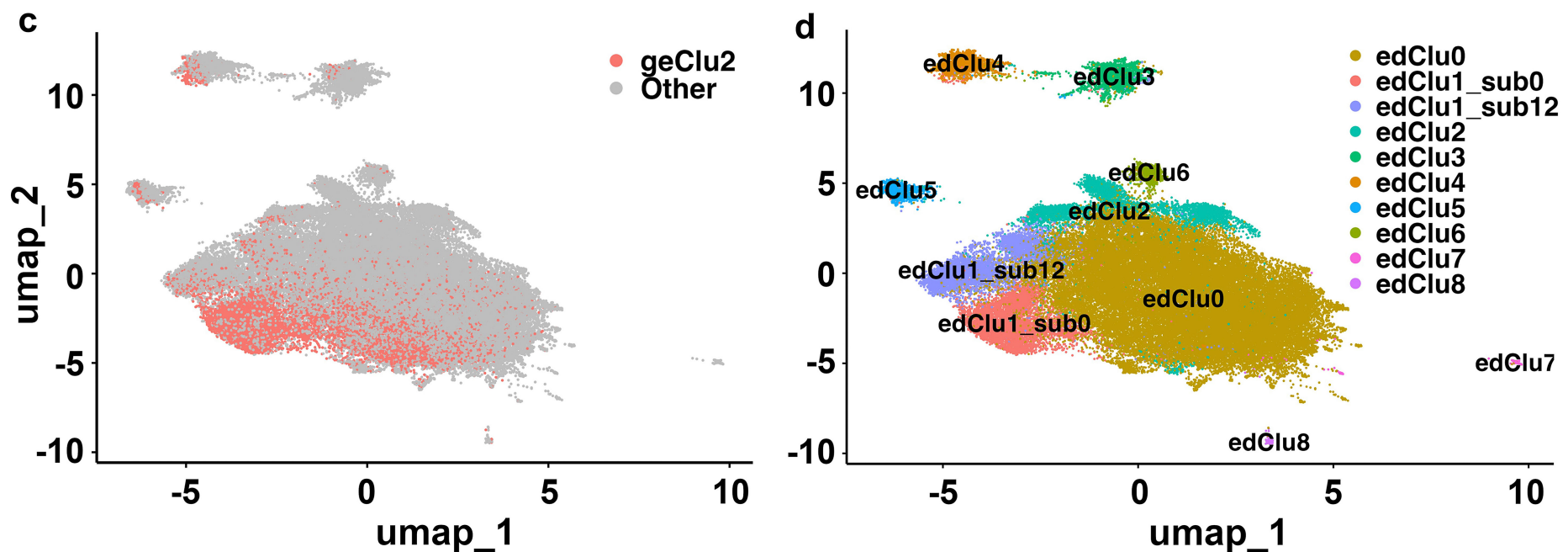
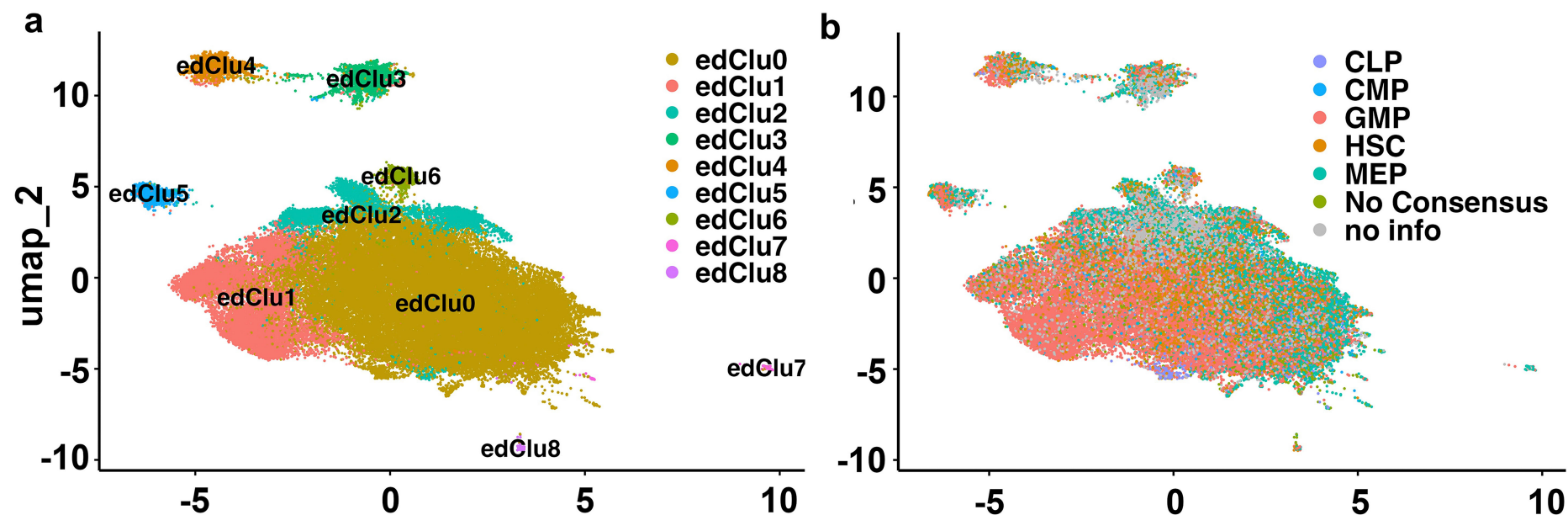
1. Marando L, Csizmar CM, Patnaik MM. Chronic myelomonocytic leukemia: molecular pathogenesis and therapeutic innovations. *Haematologica*. 2025;110(1):22-36.
2. Patnaik MM, Tefferi A. Chronic myelomonocytic leukemia: 2024 update on diagnosis, risk stratification and management. *Am J Hematol*. 2024;99(6):1142-1165.
3. Nishikura K. A-to-I editing of coding and non-coding RNAs by ADARs. *Nat Rev Mol Cell Biol*. 2016;17(2):83-96.
4. Teoh PJ, Koh MY, Chng WJ. ADARs, RNA editing and more in hematological malignancies. *Leukemia*. 2021;35(2):346-359.
5. Check Hayden E. RNA editing study under intense scrutiny. *Nature*. 2012. <https://www.nature.com/articles/nature.2012.10217> . Accessed on August 21, 2026.
6. Gu T, Buaas FW, Simons AK, Ackert-Bicknell CL, Braun RE, Hibbs MA. Canonical A-to-I and C-to-U RNA editing is enriched at 3'UTRs and microRNA target sites in multiple mouse tissues. *PLoS One*. 2012;7(3):e33720.
7. Gabay O, Shoshan Y, Kopel E, et al. Landscape of adenosine-to-inosine RNA recoding across human tissues. *Nat Commun*. 2022;13(1):1184.
8. Ferrall-Fairbanks MC, Dhawan A, Johnson B, et al. Progenitor hierarchy of chronic myelomonocytic leukemia identifies inflammatory monocytic-biased trajectory linked to worse outcomes. *Blood Cancer Discov*. 2022;3(6):536-553.
9. Mansi L, Tangaro MA, Lo Giudice C, et al. REDlportal: millions of novel A-to-I RNA editing events from thousands of RNAseq experiments. *Nucleic Acids Res*. 2021;49(D1):D1012-D1019.
10. Gu T, Fu AQ, Bolt MJ, White KP. Clinical Relevance of noncoding adenosine-to-inosine RNA editing in multiple human cancers. *JCO Clin Cancer Inform*. 2019;3:1-8.
11. Chan TW, Dodson JP, Arbet J, Boutros PC, Xiao X. Single-cell analysis in lung adenocarcinoma implicates RNA editing in cancer innate immunity and patient prgnosis. *Cancer Res*. 2023;83(3):374-385.
12. Meduri E, Breeze C, Marando L, Richardson SE, Huntly BJP. The RNA editing landscape in acute myeloid leukemia reveals associations with disease mutations and clinical outcome. *iScience*. 2022;25(12):105622.
13. Vromman M, Anckaert J, Bortoluzzi S, et al. Large-scale benchmarking of circRNA detection tools reveals large differences in sensitivity but not in precision. *Nat Methods*. 2023;20(8):1159-1169.
14. Montalban-Bravo G, Thongon N, Rodriguez-Sevilla JJ, et al. Targeting MCL1-driven anti-apoptotic pathways overcomes blast progression after hypomethylating agent failure in chronic myelomonocytic leukemia. *Cell Rep Med*. 2024;5(6):101585.
15. Oetjen KA, Lindblad KE, Goswami M, et al. Human bone marrow assessment by single-cell RNA sequencing, mass cytometry, and flow cytometry. *JCI Insight*. 2018;3(23):e124928.

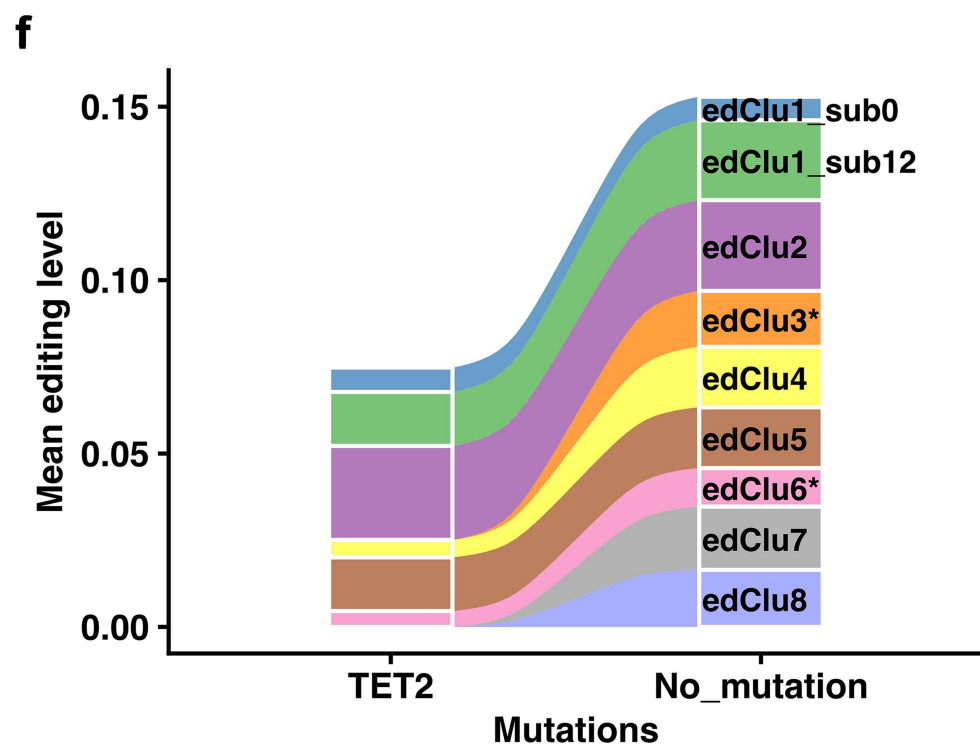
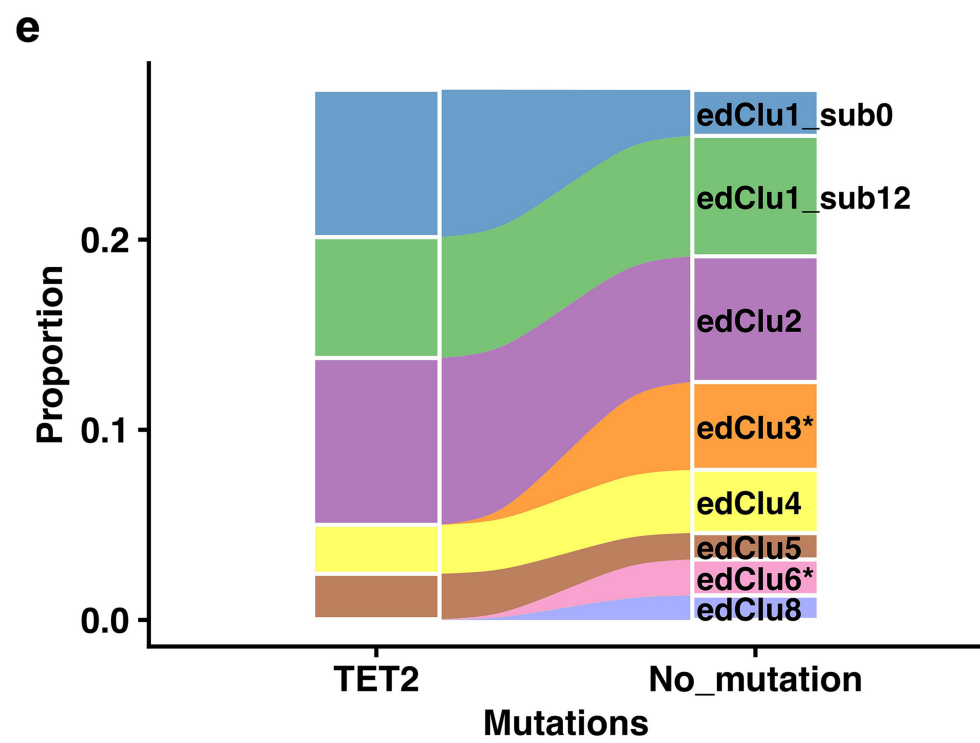
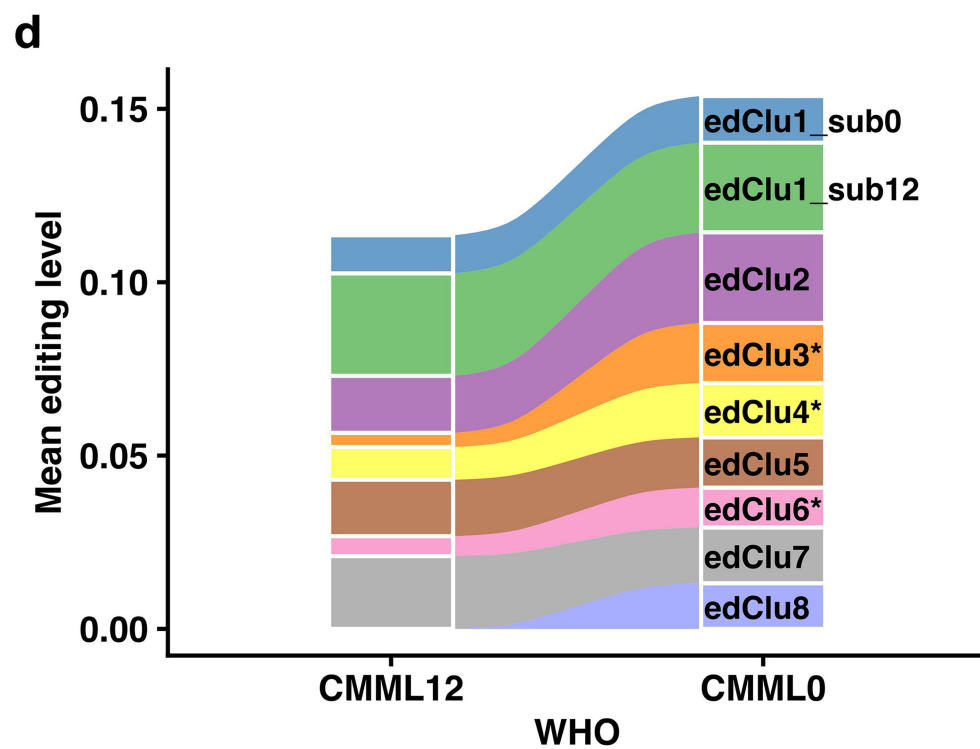
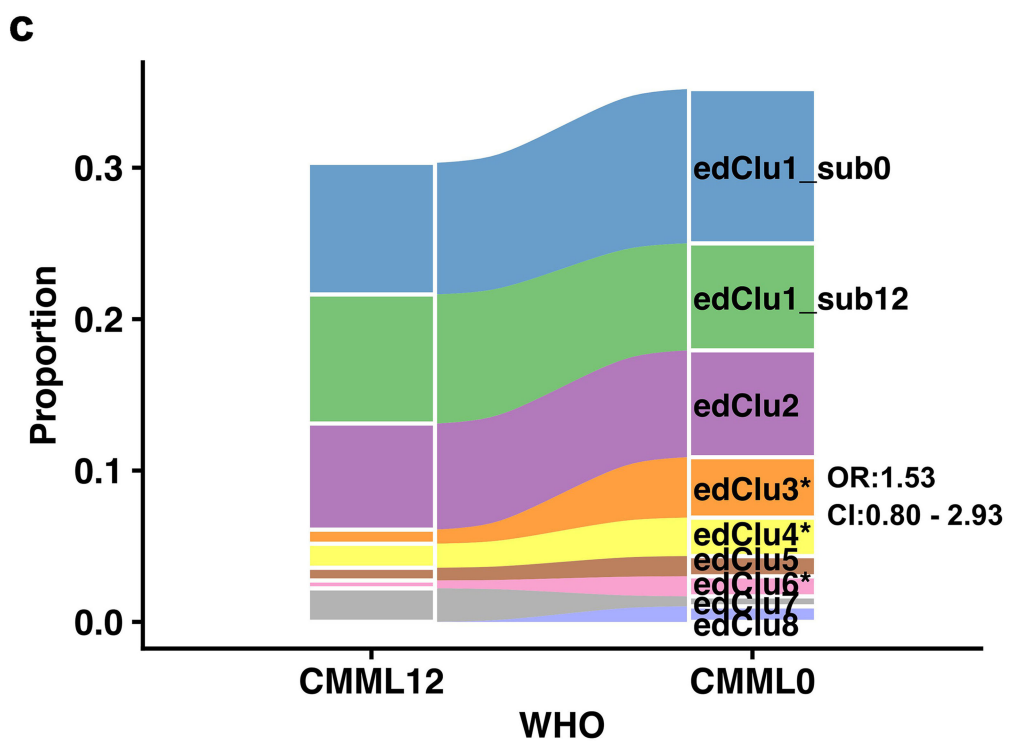
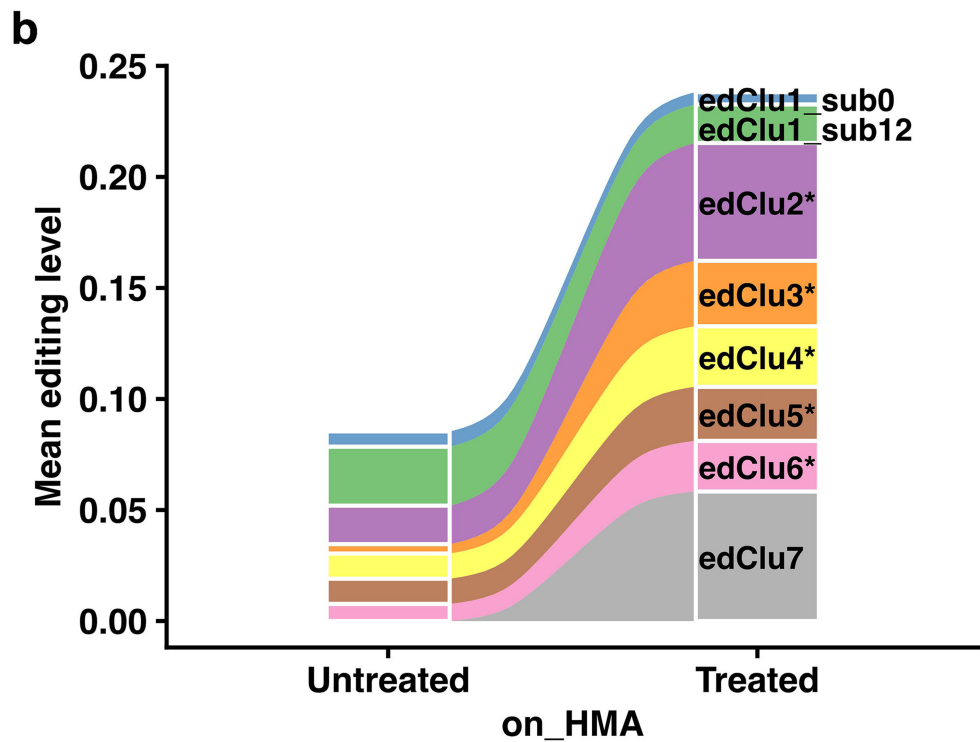
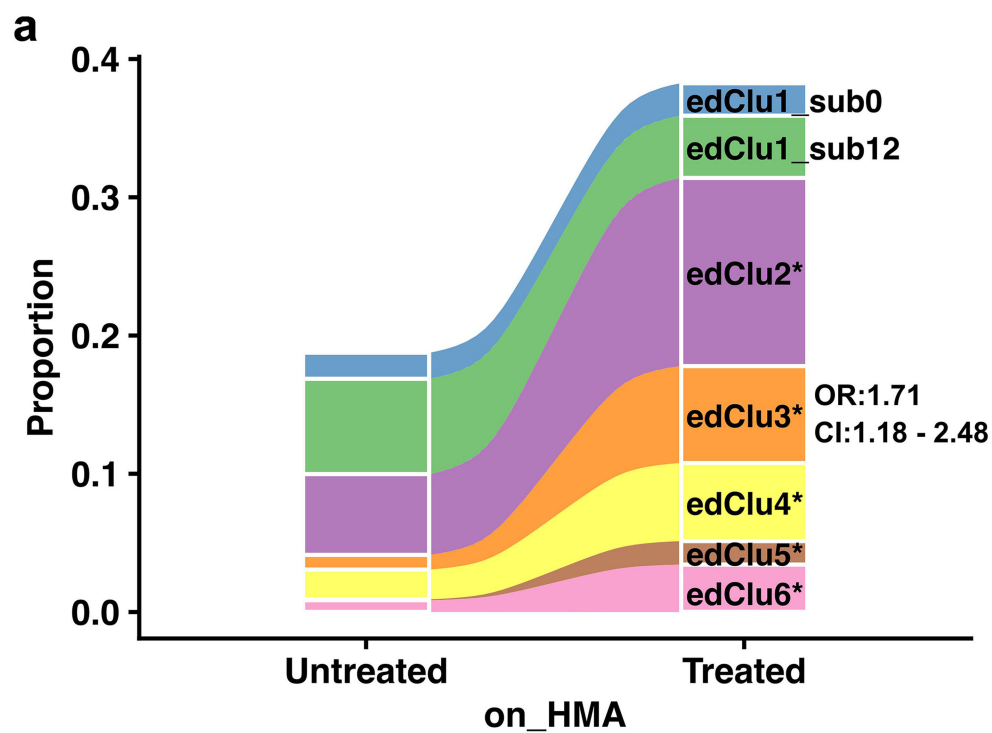
Figure legend

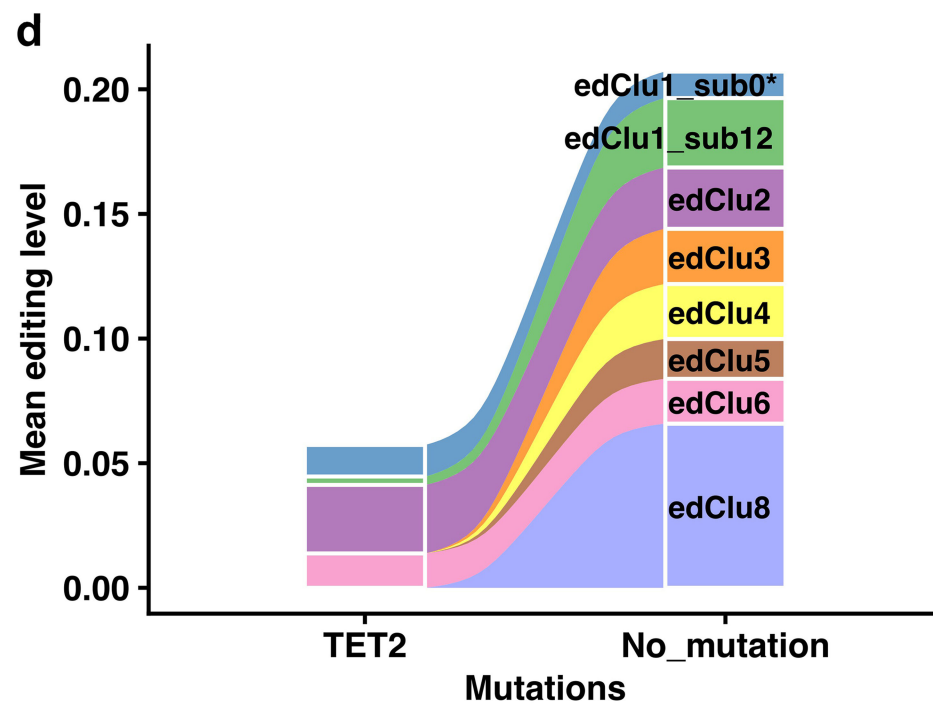
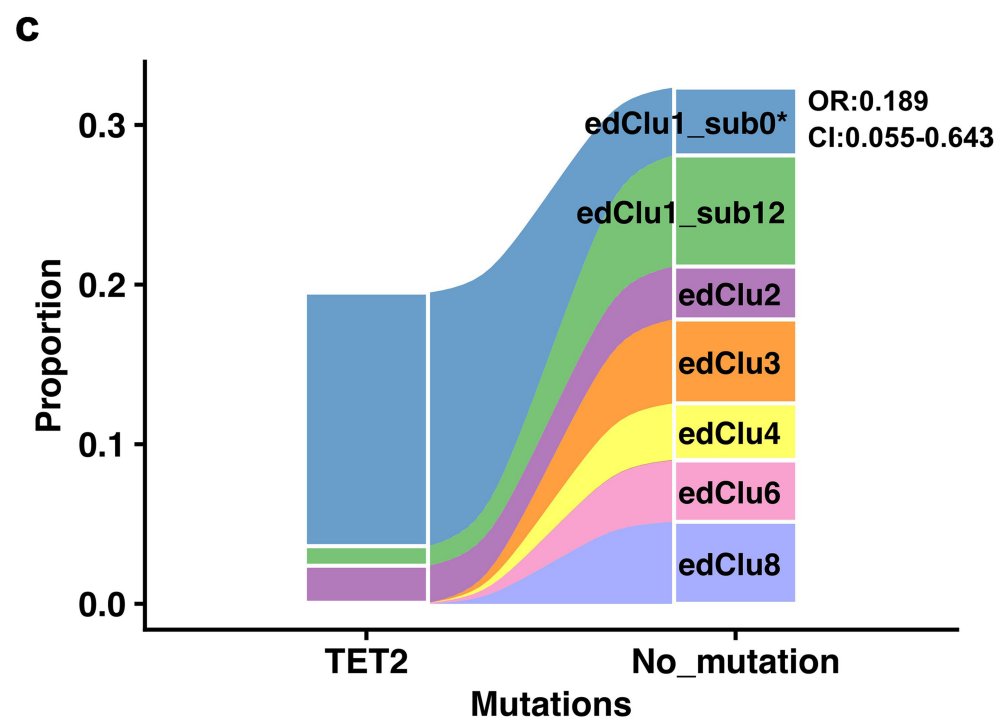
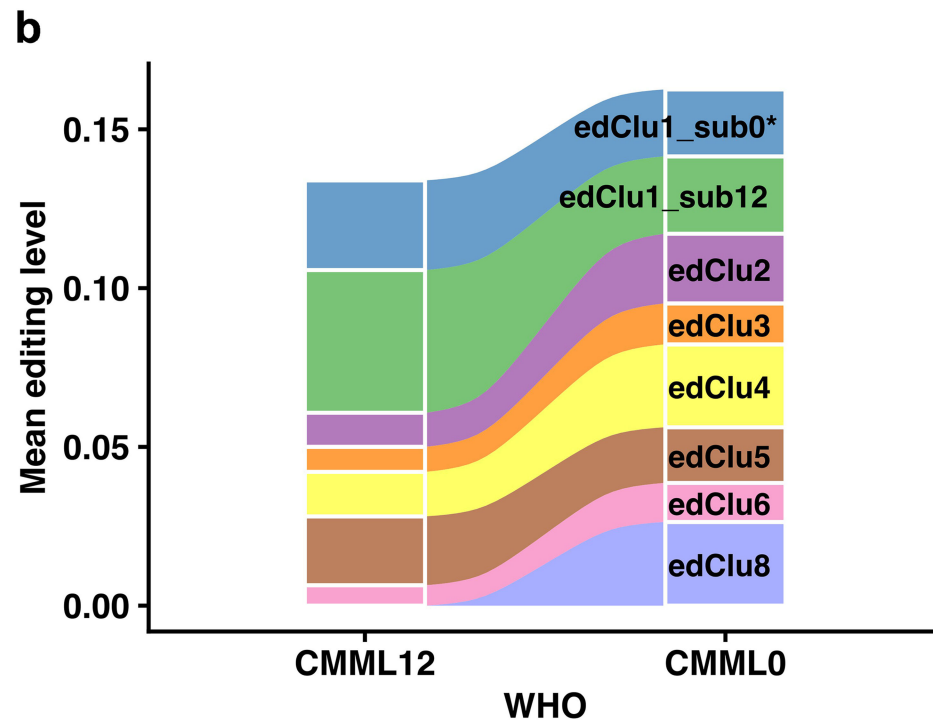
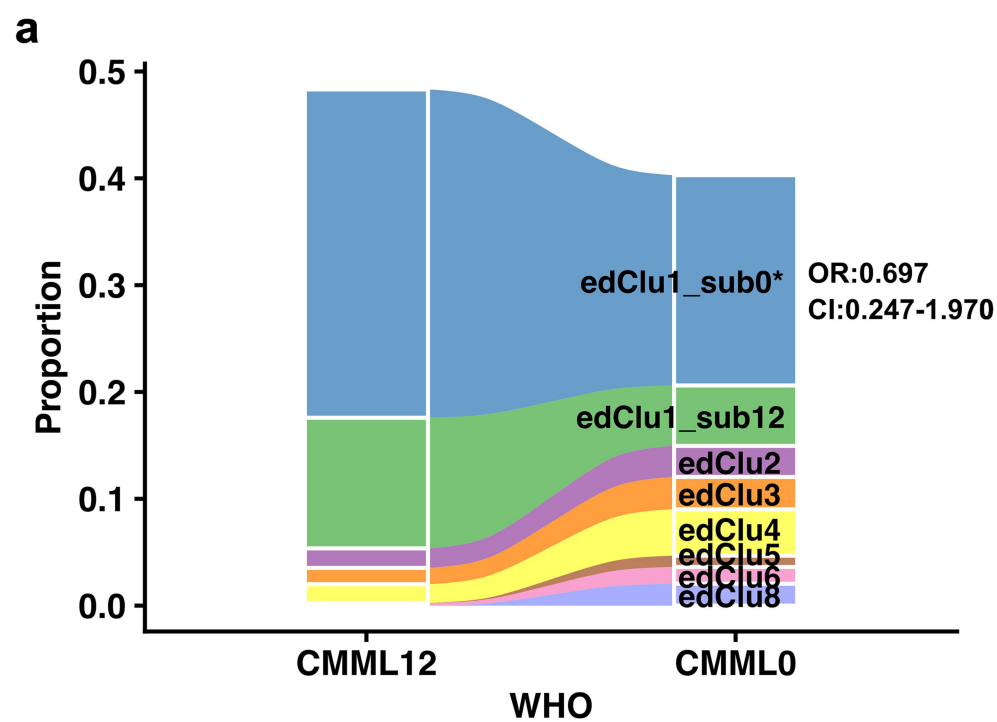
Figure 1. RNA editing-defined clusters. (a) RNA editing-defined clusters identified from single-cell RNA editing profiles. (b) Editing cells were annotated using previously reported gene expression–based cell-type identities. (c) Cells assigned to the previously reported aggressive gene expression cluster, geClu2. (d) RNA editing-defined clusters identified after subclustering of edClu1. (e) Scaled and normalized average expression and the percentage of cells expressing previously reported geClu2 marker genes across RNA editing-defined clusters. (f) Average RNA editing ratios for editing sites located within geClu2 marker genes across RNA editing-defined clusters.

Figure 2. Alterations in RNA editing-defined clusters, edClu3 and edClu6, across clinical features. (a,c,e) Association between clinical features and the proportion of cells across editing-defined clusters. For each sample, the proportion of cells assigned to each editing cluster was calculated as the number of cells in the cluster divided by the total number of cells in the sample. (b,d,f) Association between mean editing level (MEL) and clinical features. MEL was calculated for each editing site by averaging the editing ratios of all cells within each editing cluster and sample. These values were then averaged across all marker editing sites for each cluster. * denotes clusters exhibiting apparent changes. To enhance the visualization of smaller subpopulations, the largest cluster (edClu0) was excluded from this plot to prevent compression of the remaining clusters. Odds ratios (ORs) and 95% confidence intervals (CIs) are shown for edClu3 in panels a and c. For the mutation analysis in panel e, the effect size could not be estimated because no edClu3-positive samples were observed in the mutation group.

Figure 3. Alterations in RNA editing-defined cluster, edClu1_sub0, across clinical features. (a,c) Association between clinical features and the proportion of cells across editing-defined clusters. ORs and 95% CIs are shown for edClu1_sub0 in panels a and c. (b,d) Association between mean editing level (MEL) and clinical features. * denotes clusters exhibiting apparent changes. To enhance the visualization of smaller subpopulations, the largest cluster (edClu0) was excluded from this plot to prevent compression of the remaining clusters. (e) Summary table of the key RNA editing-defined clusters, including their cellular identities, RNA editing features, and associated clinical outcomes.

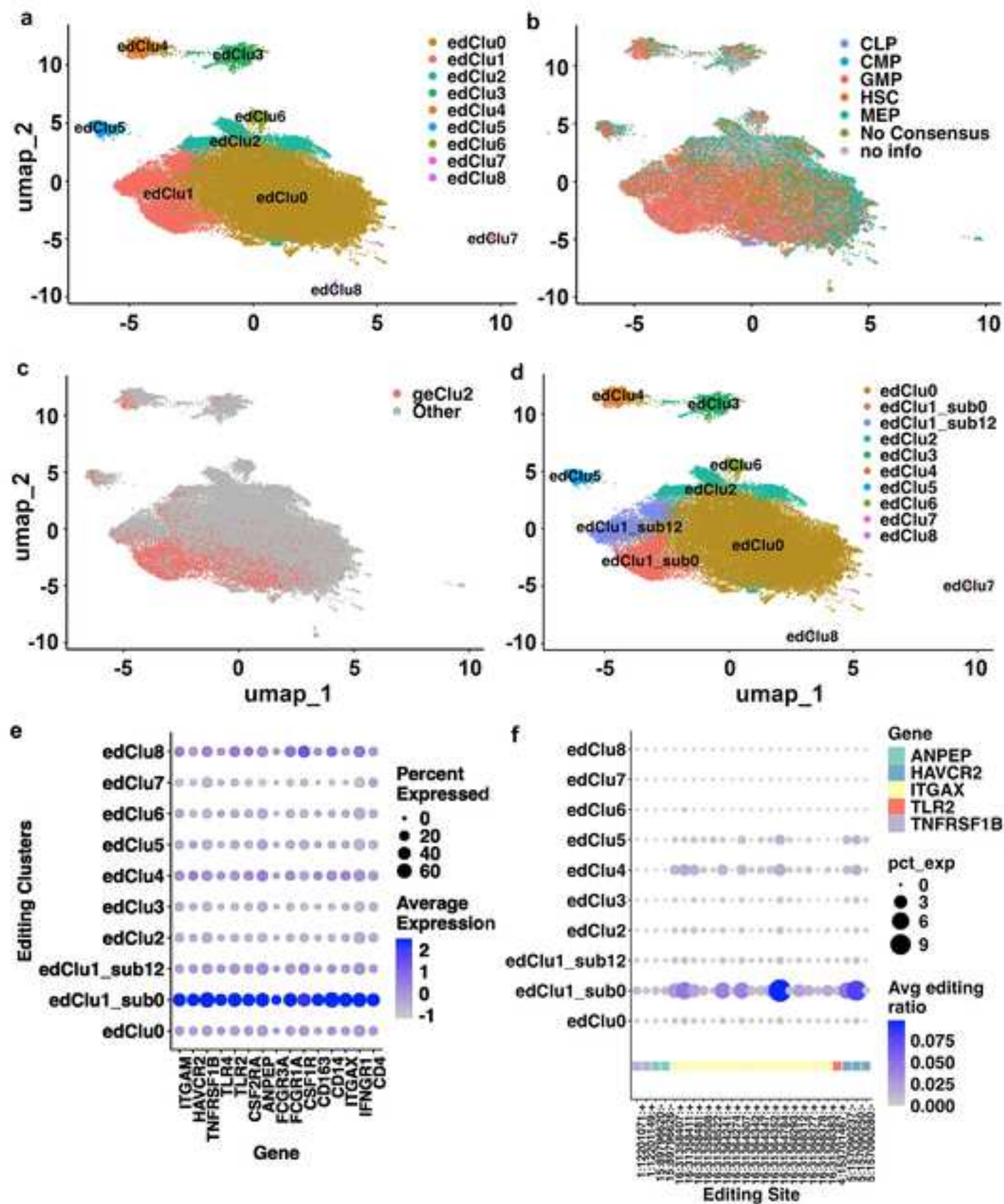


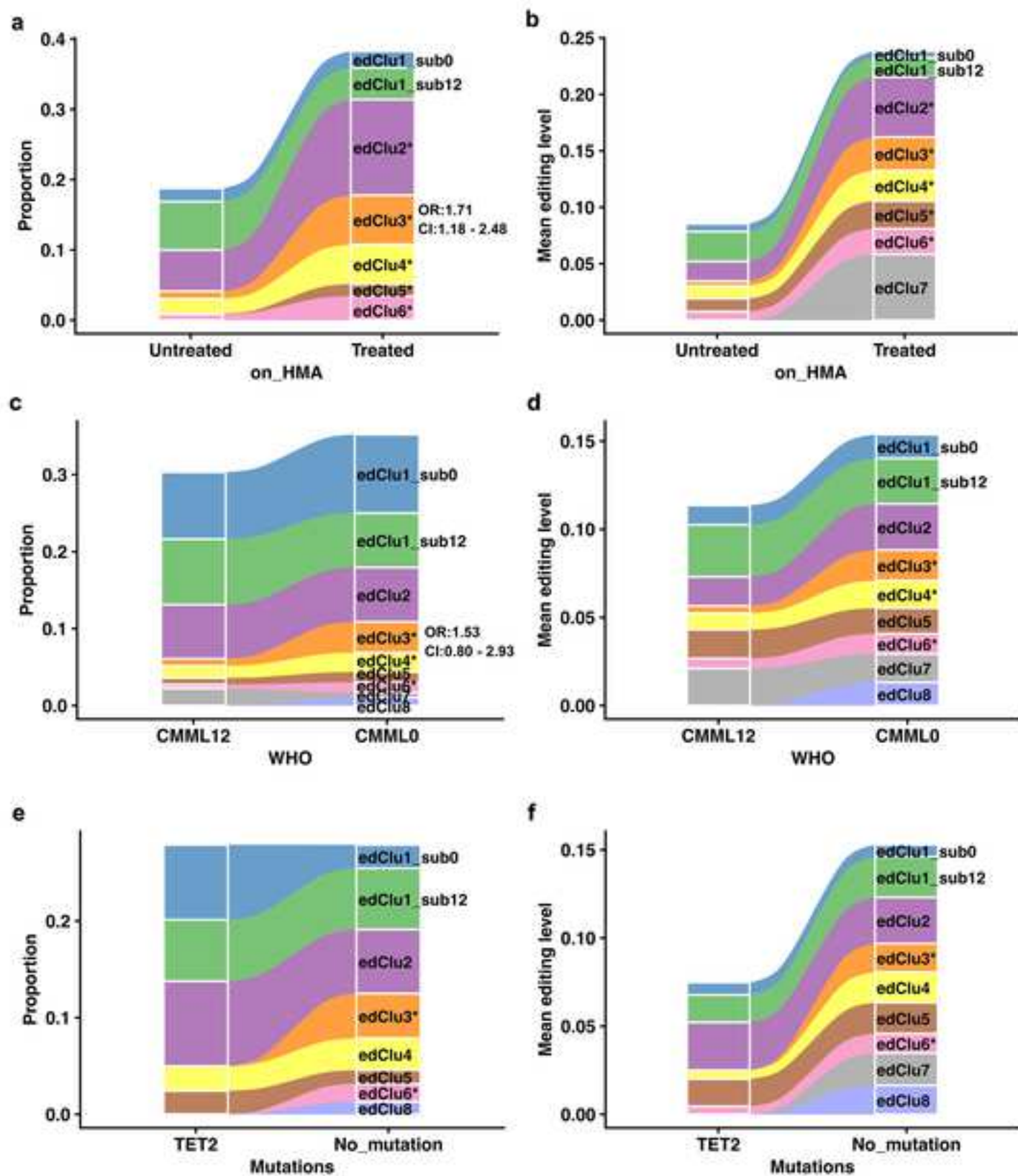


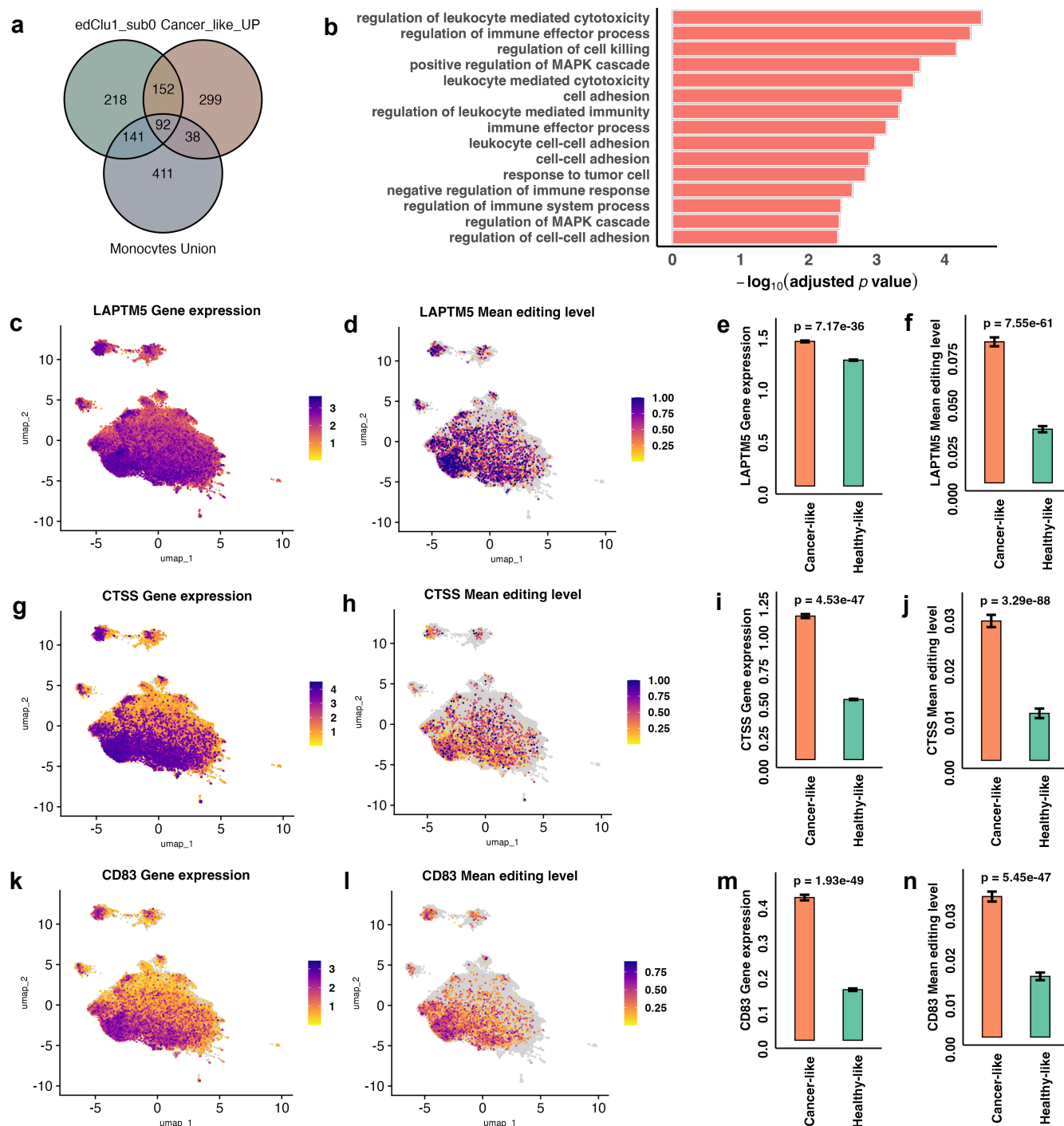


e

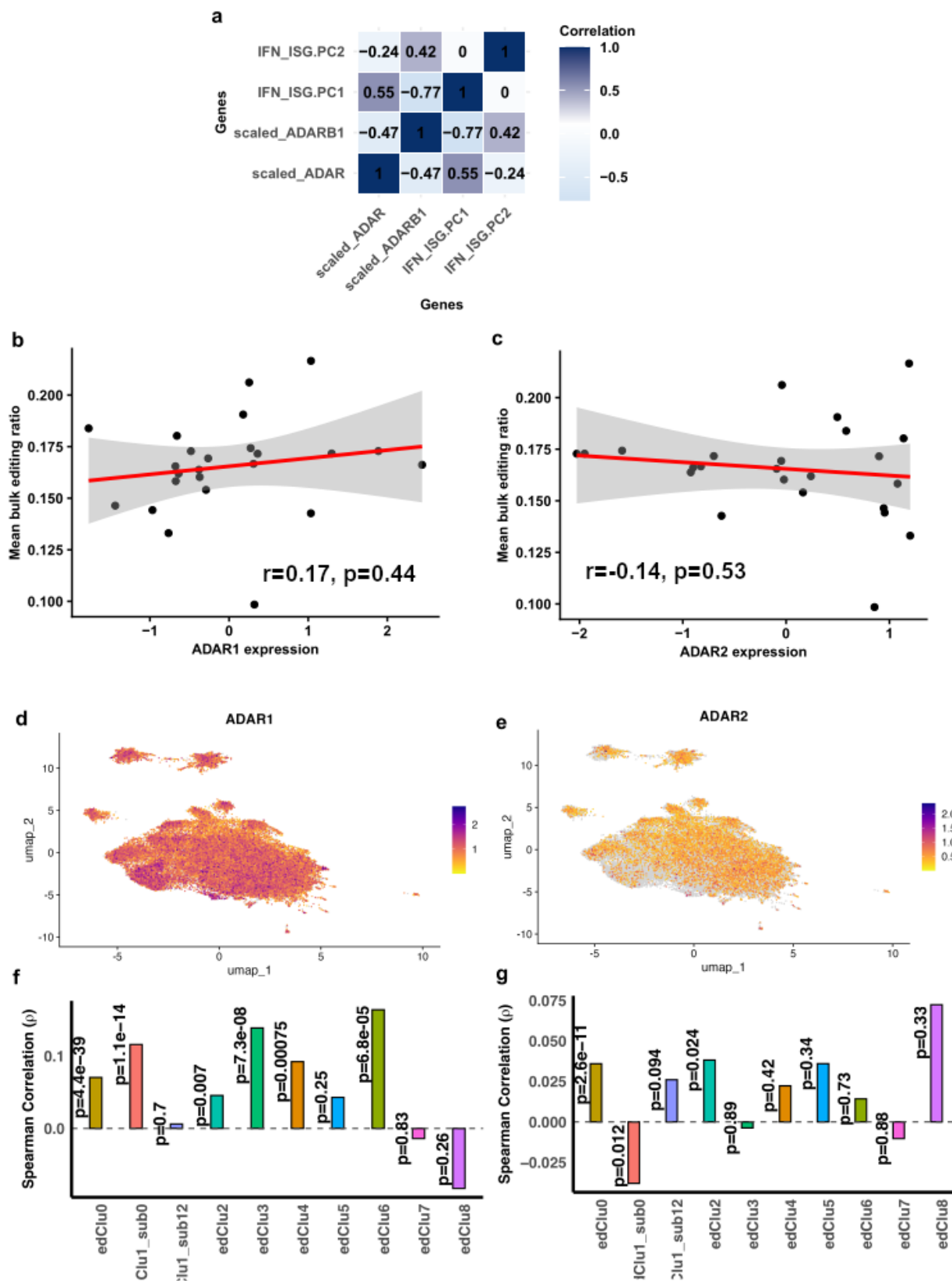
Editing-defined state	Cellular state	RNA editing characteristics	Clinical implication
edClu1_sub0	GMP-like	ADAR1-high / ADAR2-low	Associated with CMML-1/2, monocyte-biased disease, and adverse clinical outcomes
edClu3/edClu6	MEP-like	Distinct favorable editing programs	Associated with CMML-0, response to HMA therapy, and improved clinical outcomes



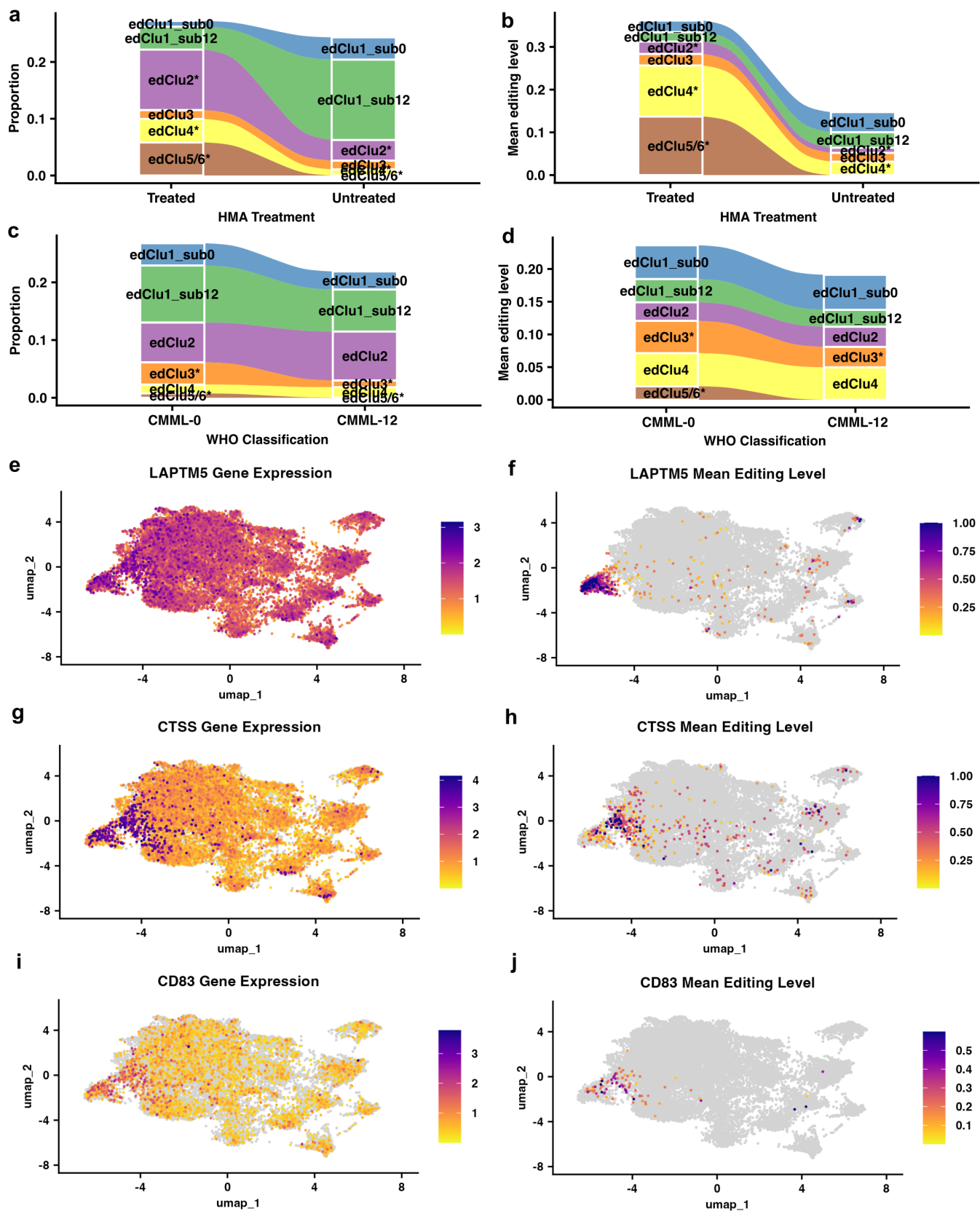




Online Supplementary Figure 1. Identification of high confident oncogenic editing sites. (a) Three approaches were used to identify high-confidence, GMP-specific oncogenic editing sites: (1) positive association with monocyte absolute count or percentage; (2) enrichment in edClu1_sub0, the most clinically aggressive GMP-associated editing state; (3) cancer-specific editing sites or sites with significant upregulation in cancer-like cells. (b) Enriched functions for the genes with high-confidence, GMP-specific oncogenic editing sites. (c,j,k) Gene expression profile of the genes of interest. (d,h,l) Mean editing levels of the genes of interest. (e,i,m) Differential expression of the genes of interest between healthy-like and cancer-like cells. (f,j,n) Differential mean editing levels of the genes of interest between healthy-like and cancer-like cells.



Online Supplementary Figure 2. *ADAR1* and *ADAR2* expression patterns and their association with RNA editing across RNA editing-defined clusters. (a) Correlation of *ADAR1* and *ADAR2* expression at bulk level. IFN_ISG_PC1 and IFN_ISG_PC2 represent the top two PCs derived from 127 interferon-related and interferon-stimulated genes (ISGs). (b-c) Correlation of *ADAR1* and *ADAR2* expression with mean bulk editing ratio. (d, e) Distribution of *ADAR1* and *ADAR2* expression at single-cell level. (f, g) Spearman correlation between *ADAR1* (left) or *ADAR2* (right) expression and MEL across editing-defined clusters. MEL is calculated across all editing sites.



Online Supplementary Figure 3. Validation of changes in RNA editing-defined clusters across clinical features and oncogenic editing sites. (a,c) Association between clinical features and the proportion of cells across editing-defined clusters. For each sample, the proportion of cells assigned to each editing cluster was calculated as the number of cells in the cluster divided by the total number of cells in the sample. (b,d) Association between MEL and clinical features. MEL was computed for each editing site by averaging the editing ratios of all cells within each editing cluster and sample. These values were then averaged across all marker editing sites for each cluster. (e,g,i) Gene expression profile of the gene of interest. (f,h,j) Mean editing levels of the gene of interest. * denotes clusters exhibiting apparent changes.