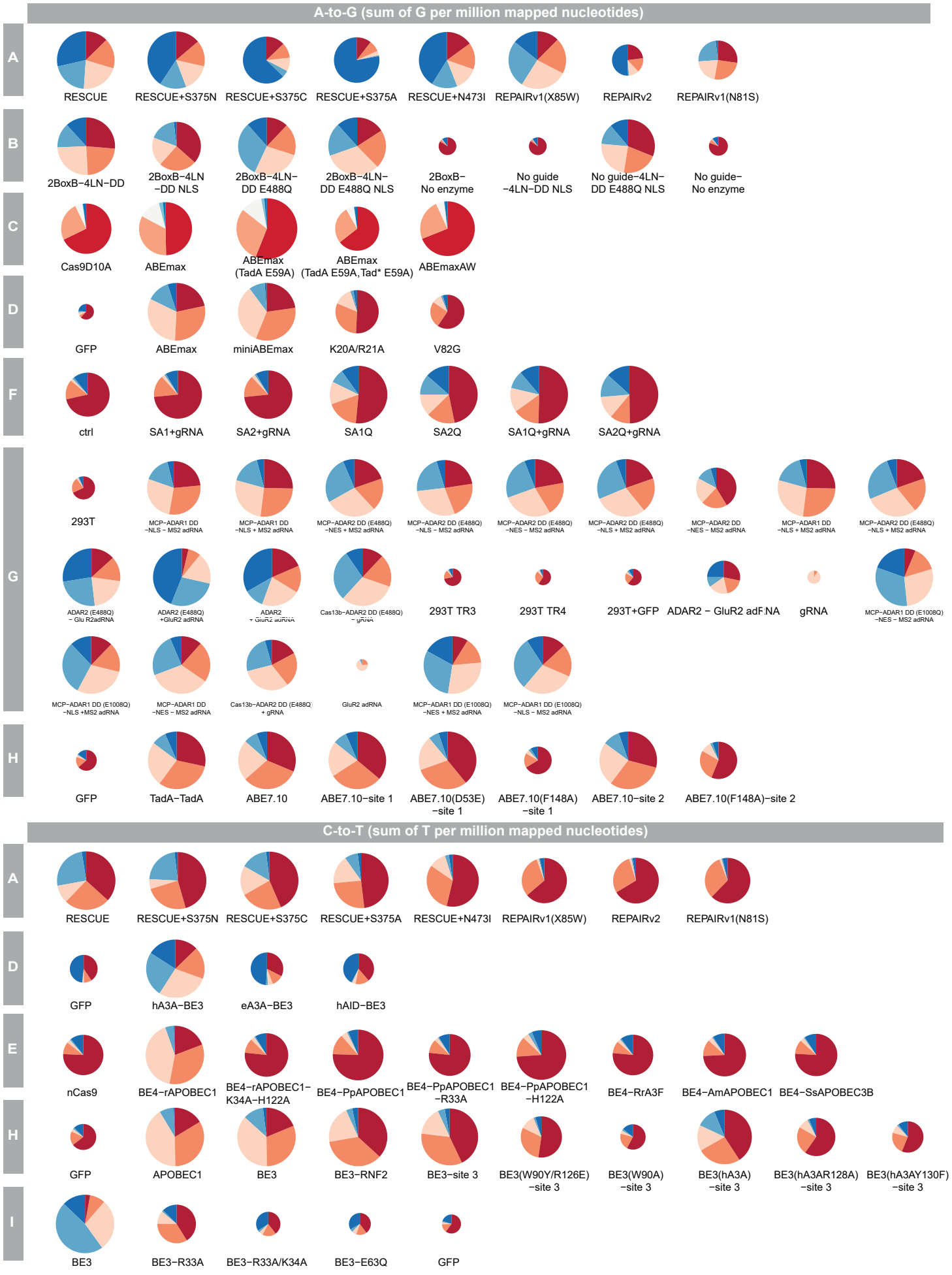
Supplementary Information to the paper:

Global quantification exposes abundant low-level off-target activity by base editors

Ilana Buchumenski[1], Shalom Hillel Roth[1], Eli Kopel[1], Efrat Katsman[1], Ariel Feiglin[1], Erez Y. Levanon[1,2*], Eli Eisenberg[3*]

Supplemental Figures

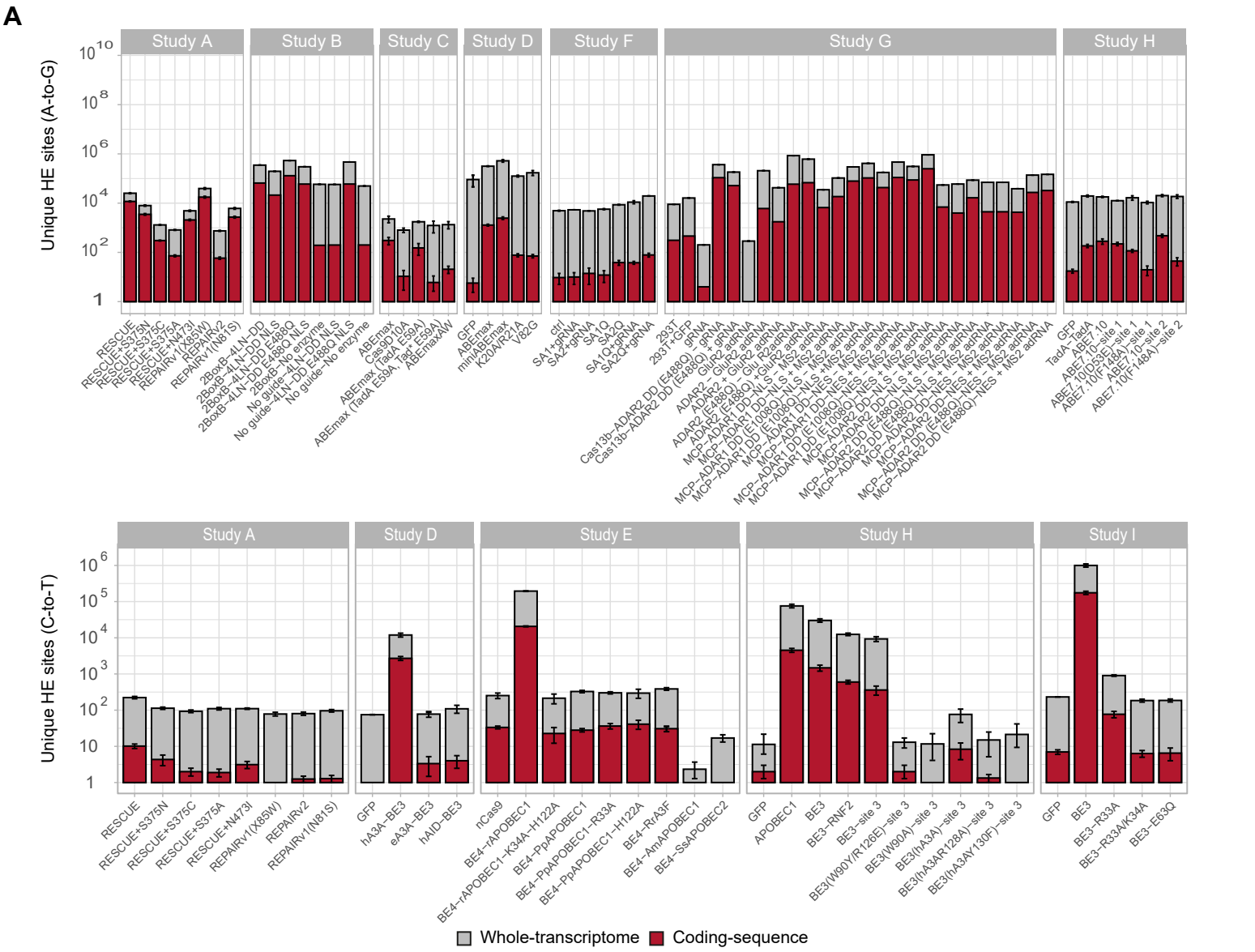
Supplemental Fig. S1

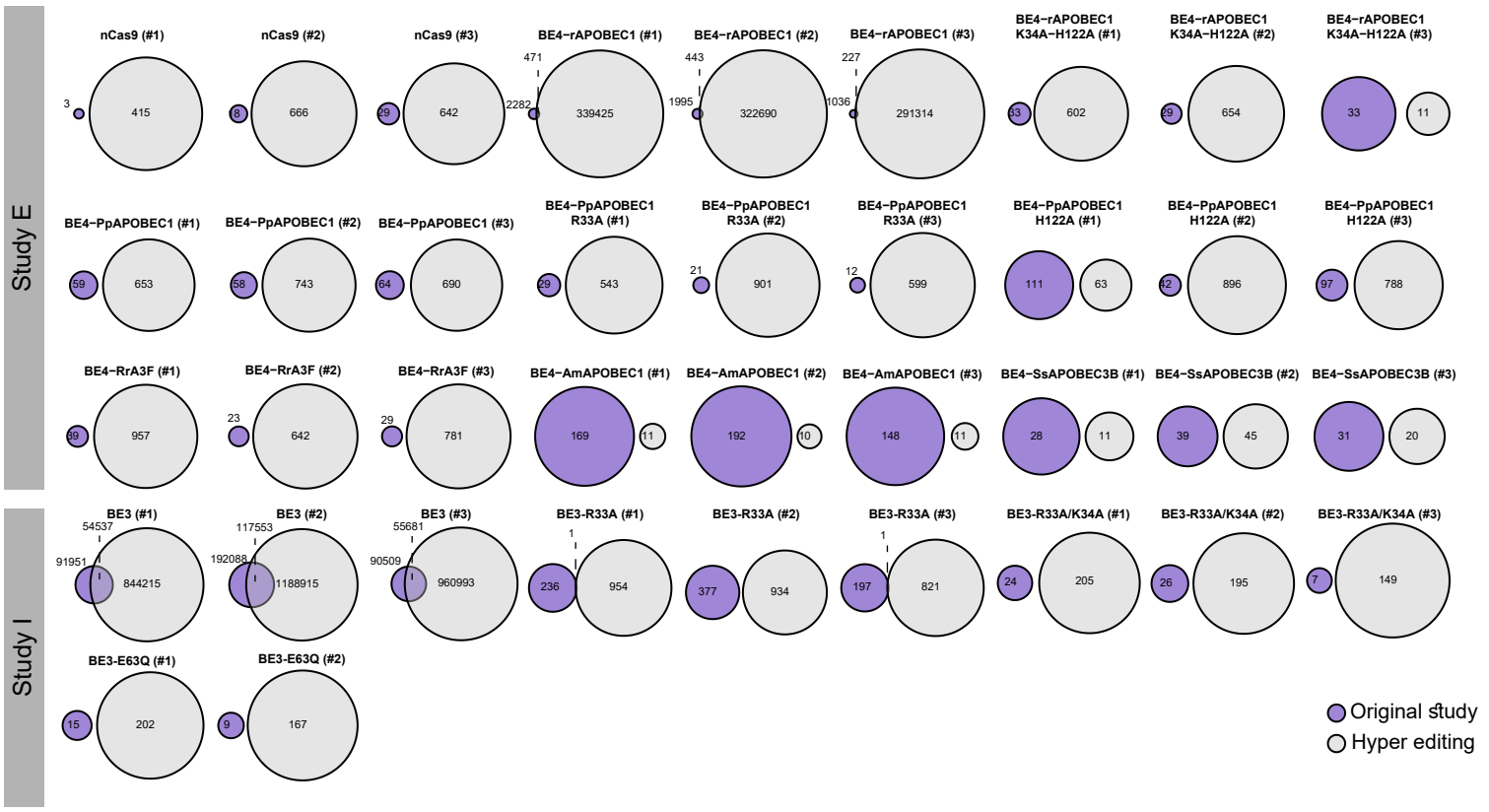


Supplemental Fig. S1. Most off-target activity occurs at weakly-edited sites.

Relative contribution of different edited bases to the editing index, for sites grouped by their editing level. To allow for an accurate editing level estimation, only positions covered by at least 50 reads were considered. Pie areas are proportional to the absolute number of detected off-target events. Replicates were pooled. The editing events observed in the controls correspond to the endogenous editing activity, mainly by the A-to-I editor ADAR.

Supplemental Fig. S2

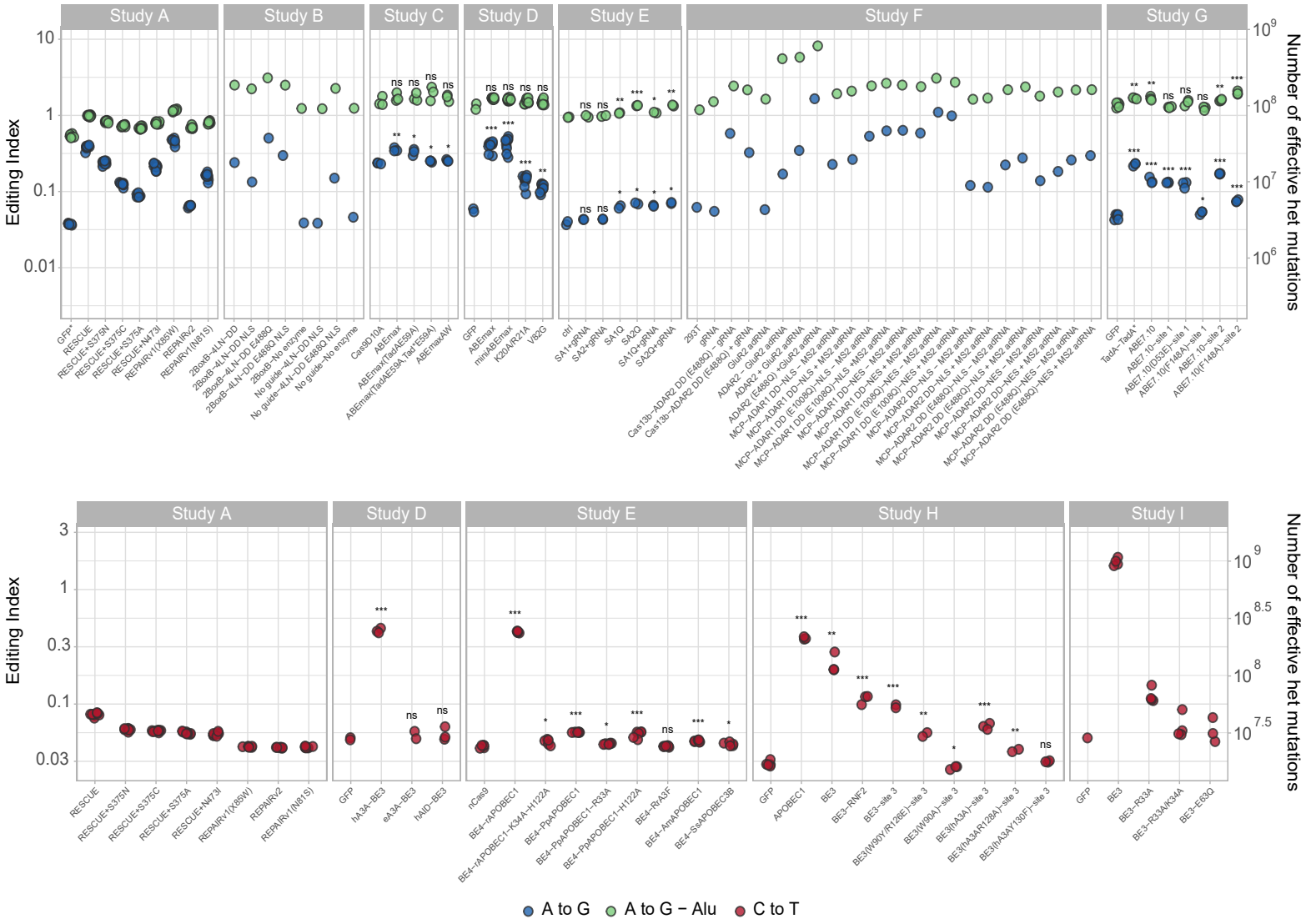




Supplemental Fig. S2. Base editors induce abundant hyper-editing.

(A) Thousands of edited sites are detected by the hyper-editing method in samples taken from eight recent studies (whole-transcriptome: grey, coding-sequence: red). These reads are not mapped by standard aligners, and their contribution was not analyzed previously. **(B)** The overlap between the off-target sites identified in original studies E and I and the sites found by the hyper-editing method. Each replicate is presented separately. In most cases these sets of sites are disjoint, suggesting that much of the non-specific off-target base-editing activity is invisible to standard genomic variation tools. The data per sample is available in Supplemental Table S6. Note that the hyper editing approach detects reads harboring large clustered of mismatches, which are not aligned to the genome using conventional mapping approaches. Therefore, one expects a low overlap between the mismatch sites found by standard tools and those identified as hyper-edited.

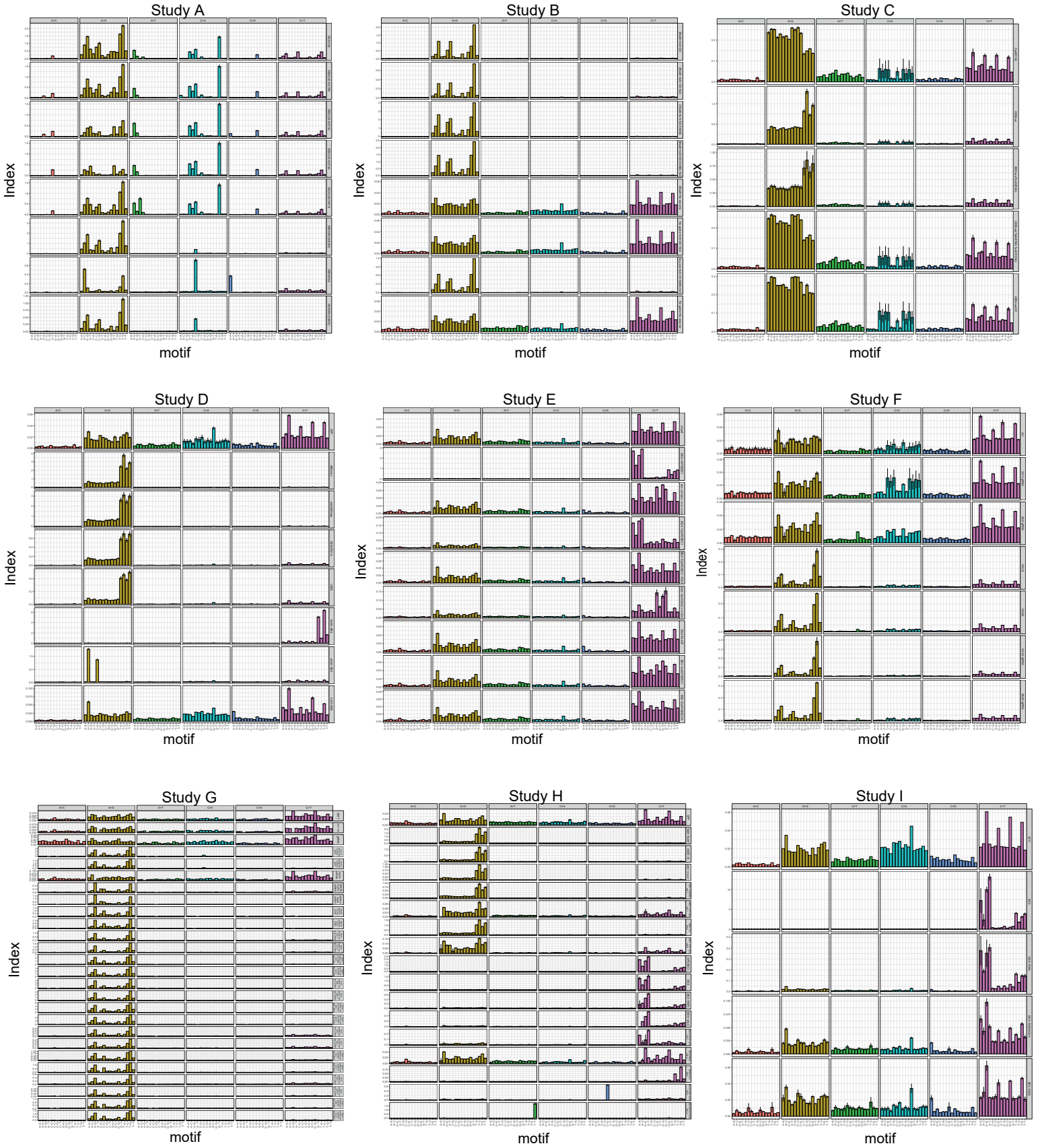
Supplemental Fig. S3



Supplemental Fig. S3. Genome-wide editing index.

(A) A-to-G index calculated over the whole genome (blue) and *Alu* elements (green), for all Adenine base-editor samples. The genome-wide index is significantly higher compared to the control samples in all studied cases (two-sided *t*-test for log(index); * p<0.05, ** p<0.01; *** p<0.001). The elevated index is equivalent to up to tens of thousands of genomic mutations (right vertical scale). The index over *Alu* elements is also substantially elevated in all cases. **(b)** C-to-T index over the whole genome, for Cytidine base-editors. Numbers are averaged over replicates, and the error-bars give the s.e.m. The data per sample is available in Supplemental Table S7. Exact p-values are presented in Supplemental Table S8.

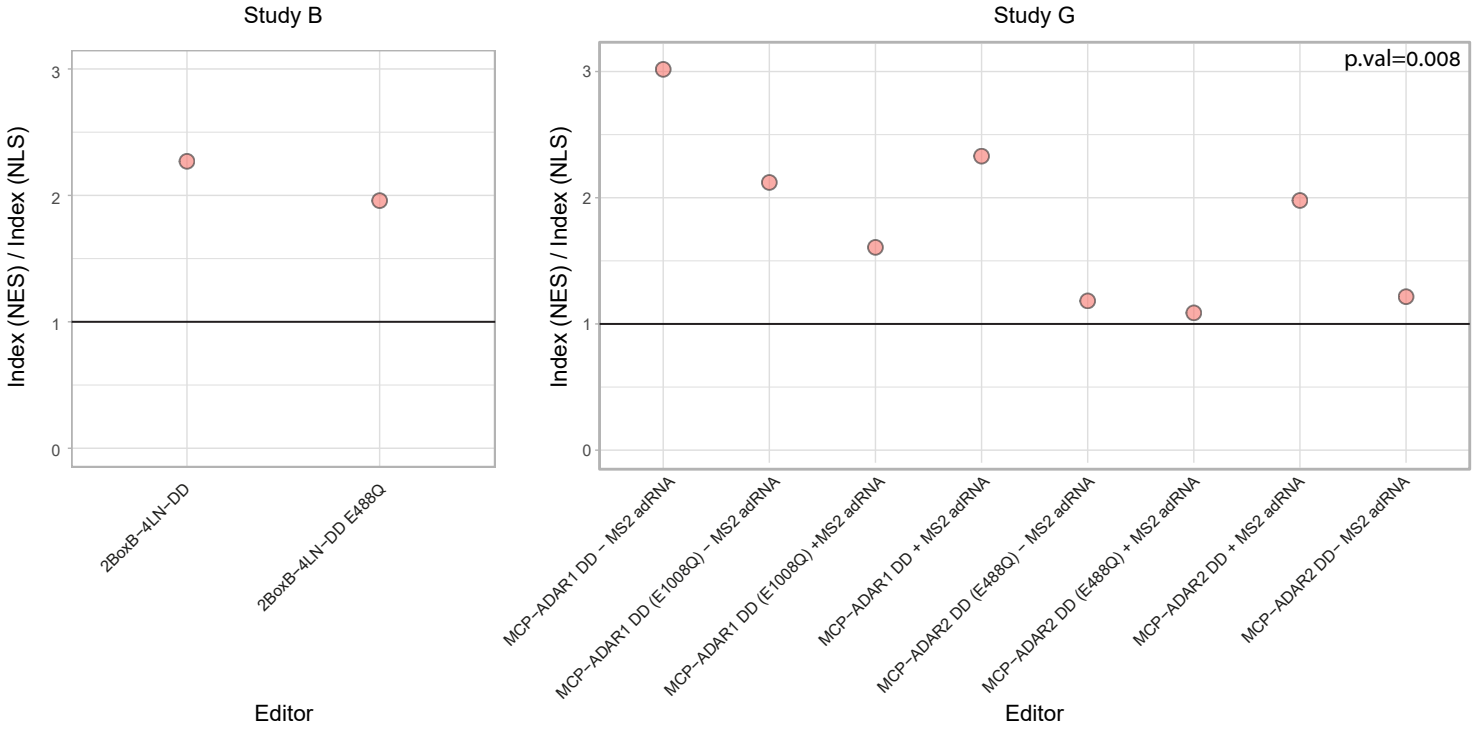
Supplemental Fig. S4



Supplemental Fig. S4. Sequence motif signature.

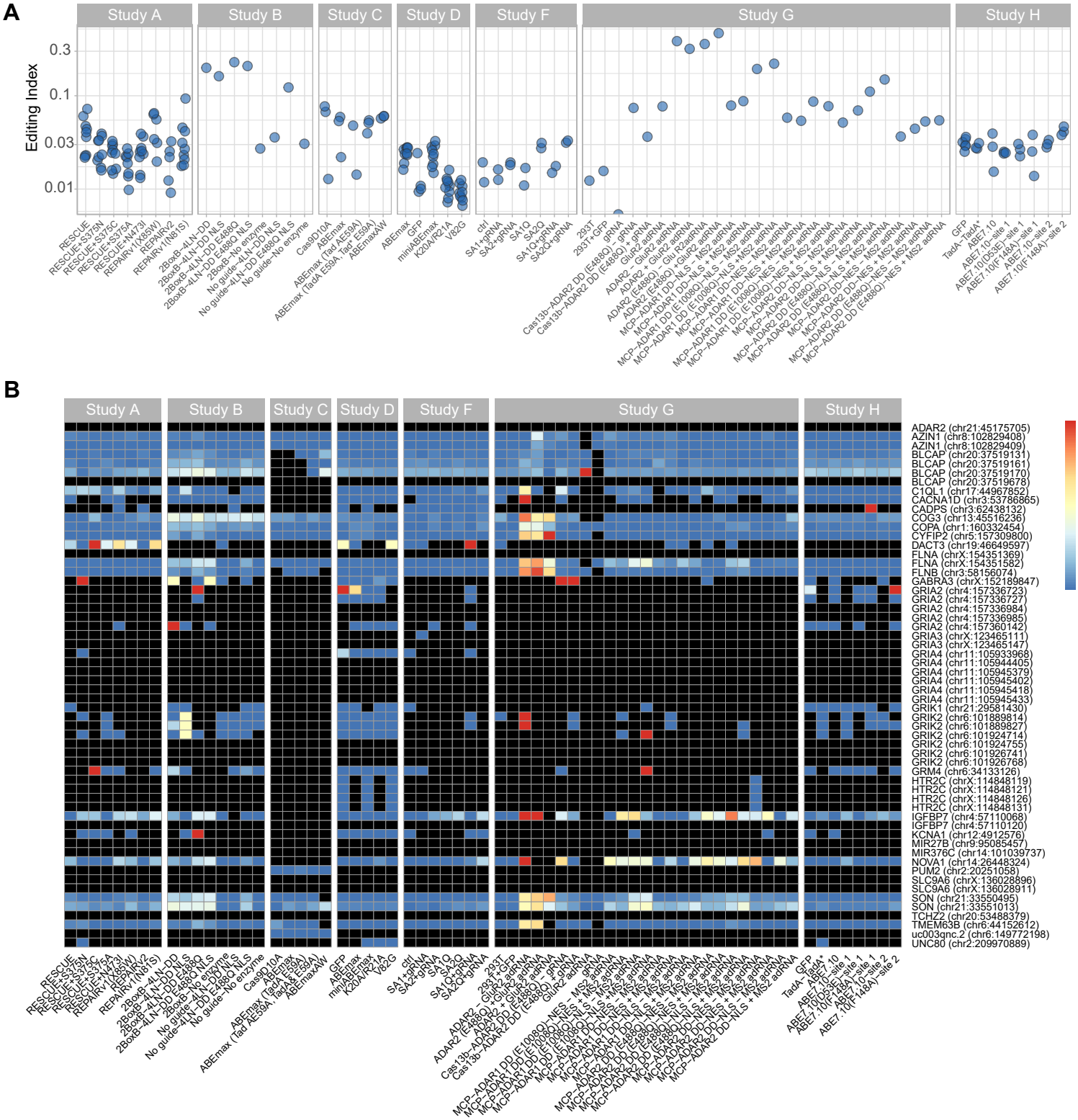
Base editors exhibit a unique sequence motif surrounding the mismatch sites, which is distinctive from that of control samples. For each base editor, the corresponding transition is marked by a black frame (A>G for A-to-G base editors, C>T for C-to-T base editors).

Supplemental Fig. S5



Supplemental Fig. S5. Nuclear localization signal affects non-specific off-target levels. Fold-increase of A-to-G index for editors containing nuclear export signal (NES) compared to a nuclear localization signal (NLS) (Studies B and G). P value was calculated using paired Wilcoxon test for the eight enzyme pairs of study G.

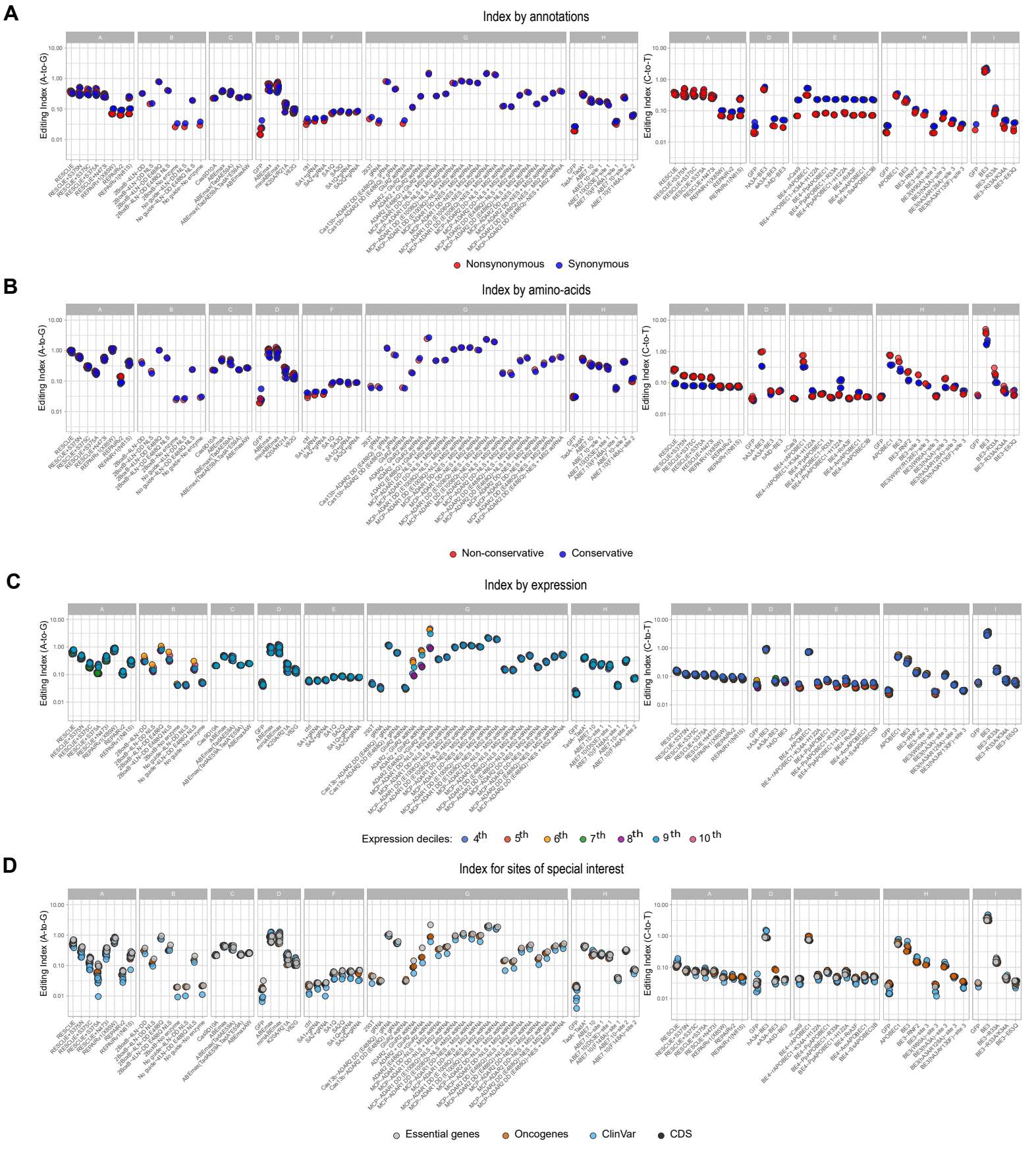
Supplemental Fig. S6



Supplemental Fig. S6. Editing at conserved A-to-I recoding sites.

(A) Total A-to-I deamination rate (weighted editing level) at endogenously edited sites within mammalian coding regions. Each circle represents one replicate. **(B)** The editing profile across all endogenously-edited, mammalian conserved, sites. Replicates were pooled. Black pixels denote low coverage (<10 reads). The data per sample is available in Supplemental Table S9.

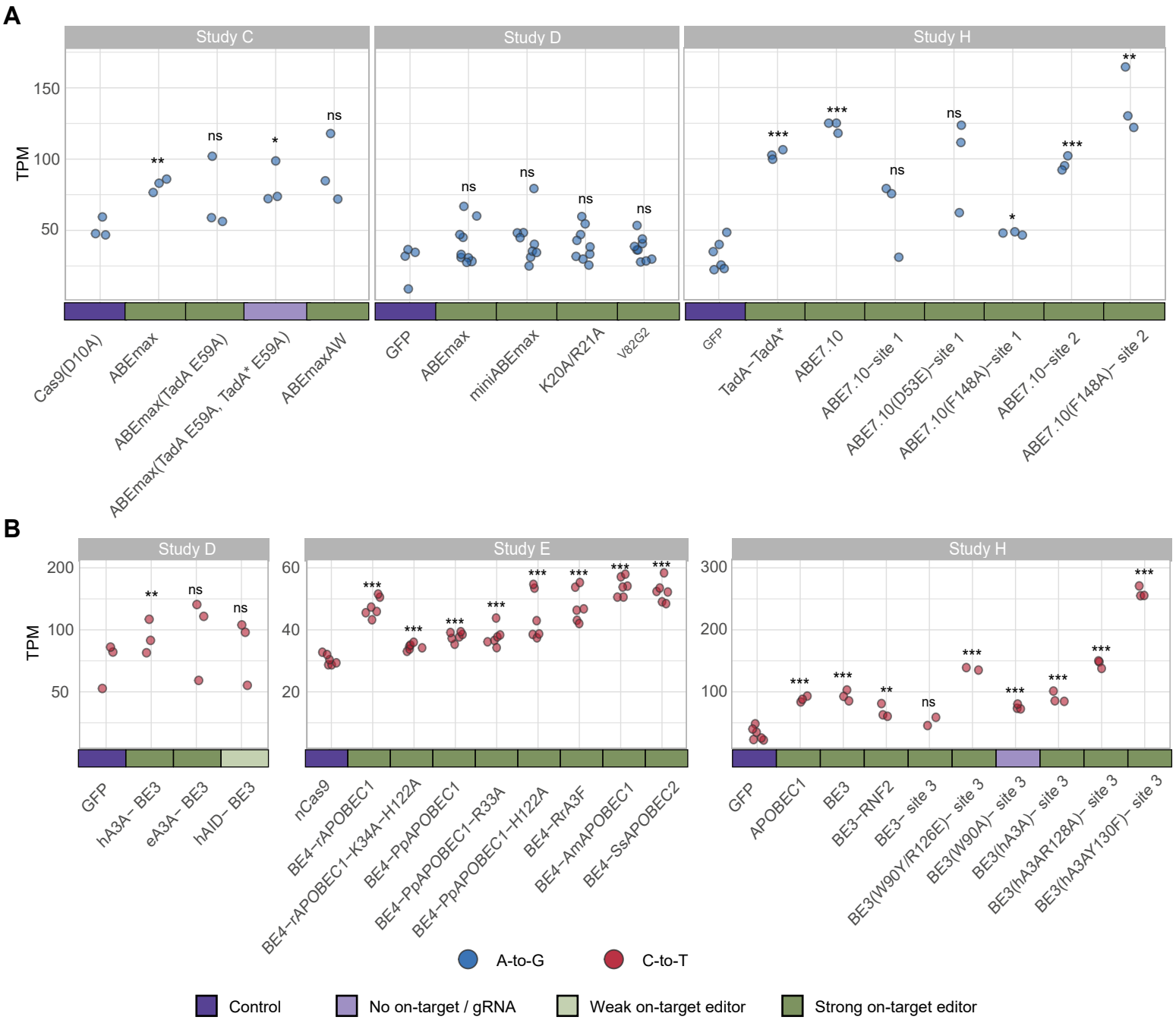
Supplemental Fig. S7



Supplemental Fig. S7. Most of the off-target activity is nonspecific.

Global off-target activity is presented for all samples studied, for **(A)** synonymous and non-synonymous (including stop-gain and stop-loss) mutations, **(B)** conservative and non-conservative amino-acid substitutions (see Methods) **(C)** broken by deciles of gene expression, **(D)** essential genes, oncogenes, and a set of clinically relevant, pathogenic, amino-acid substitutions (ClinVar, see Methods), compared with the overall index in coding sequence. Each circle represents one sample. The data per sample is available in Supplemental Table S10

Supplemental Fig. S8



Supplemental Fig. S8. Elevated expression of the apoptosis-related gene p21 upon introduction of base-editors.
CDKN1A (also known as p21) plays a critical role in the cellular response to DNA damage and overexpression of its transcript (*cdkn1a*) results in cell cycle arrest. *cdkn1a* expression levels are presented for samples expressing various base-editors and matching controls in units of transcripts per million (TPM) for (A) Adenine-deaminase base-editors (Studies C, D, H) and (B) Cytidine deaminase base-editors (Studies D, E, H,). Expression is significantly elevated for most enzymes examined (two-sided *t*-test for TPM; * p<0.05, ** p<0.01; *** p<0.001). Significance was not assessed for studies A, F and I (due to an insufficient number of control samples) and studies B and G (providing a single replicate per condition). The data per sample and exact p-values are presented in Supplemental Table S11.