SUPPLEMENTARY FIGURES

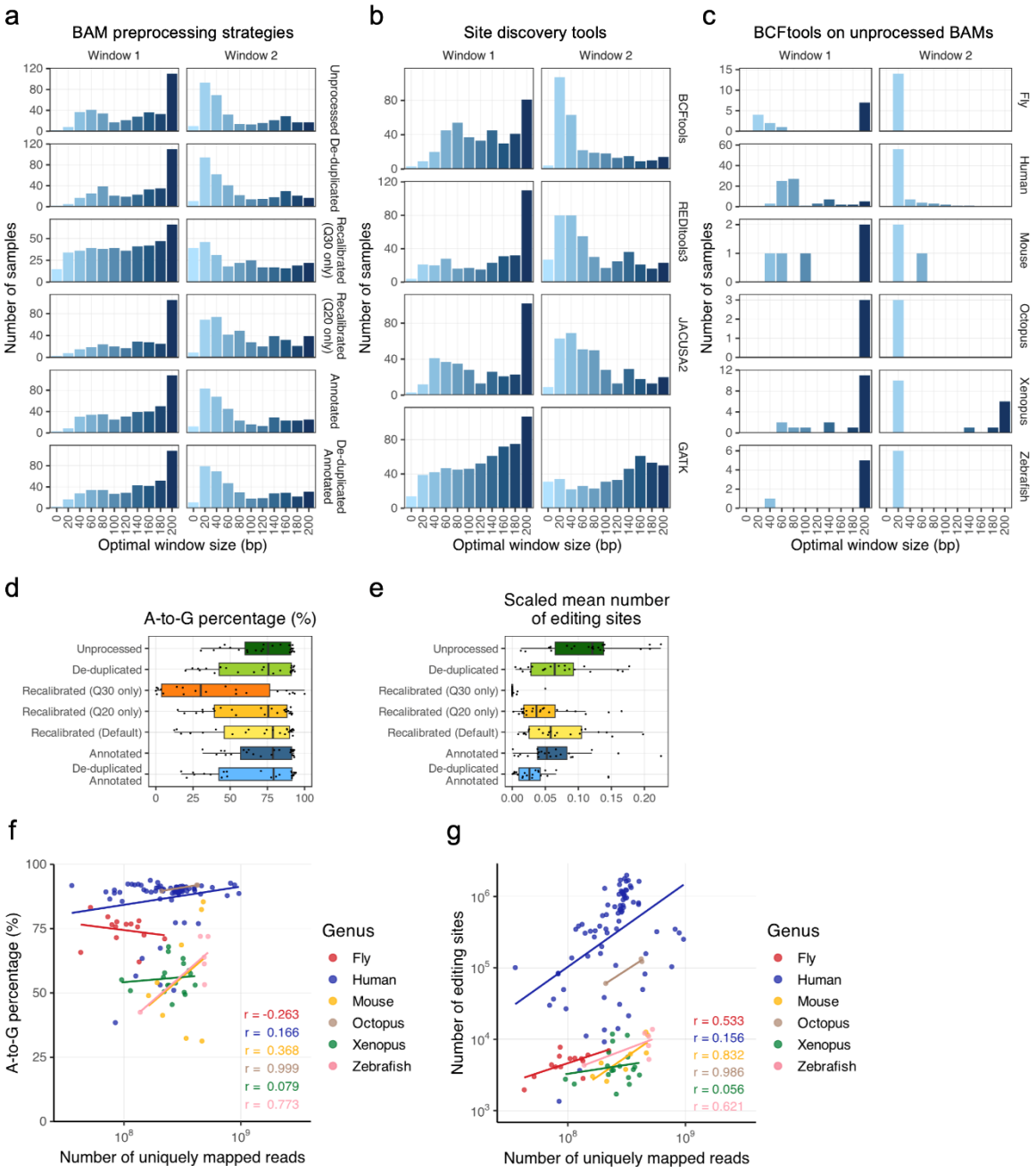

Figure S1| Additional benchmarking results for various pipeline features and sequencing depth effects.
(a-c) Optimal dynamic window sizes for **(a)** all methods summarized by BAM preprocessing strategy, **(b)** all methods summarized by discovery tool, and **(c)** comparisons between genera using BCFtools on unprocessed BAMs. **(d-e)** Box plots of key performance metrics, showing **(d)** A-to-G percentage and **(e)** scaled mean number of editing sites for GATK-only filters including “default” GATK settings. **(f-g)** Scatter plots showing performance metrics versus number of uniquely mapped reads. **(f)** Sequencing depth does not affect A-to-G percentages but **(g)** appears to have a positive trend with the number of detected sites. Pearson’s r values are indicated on the plots.

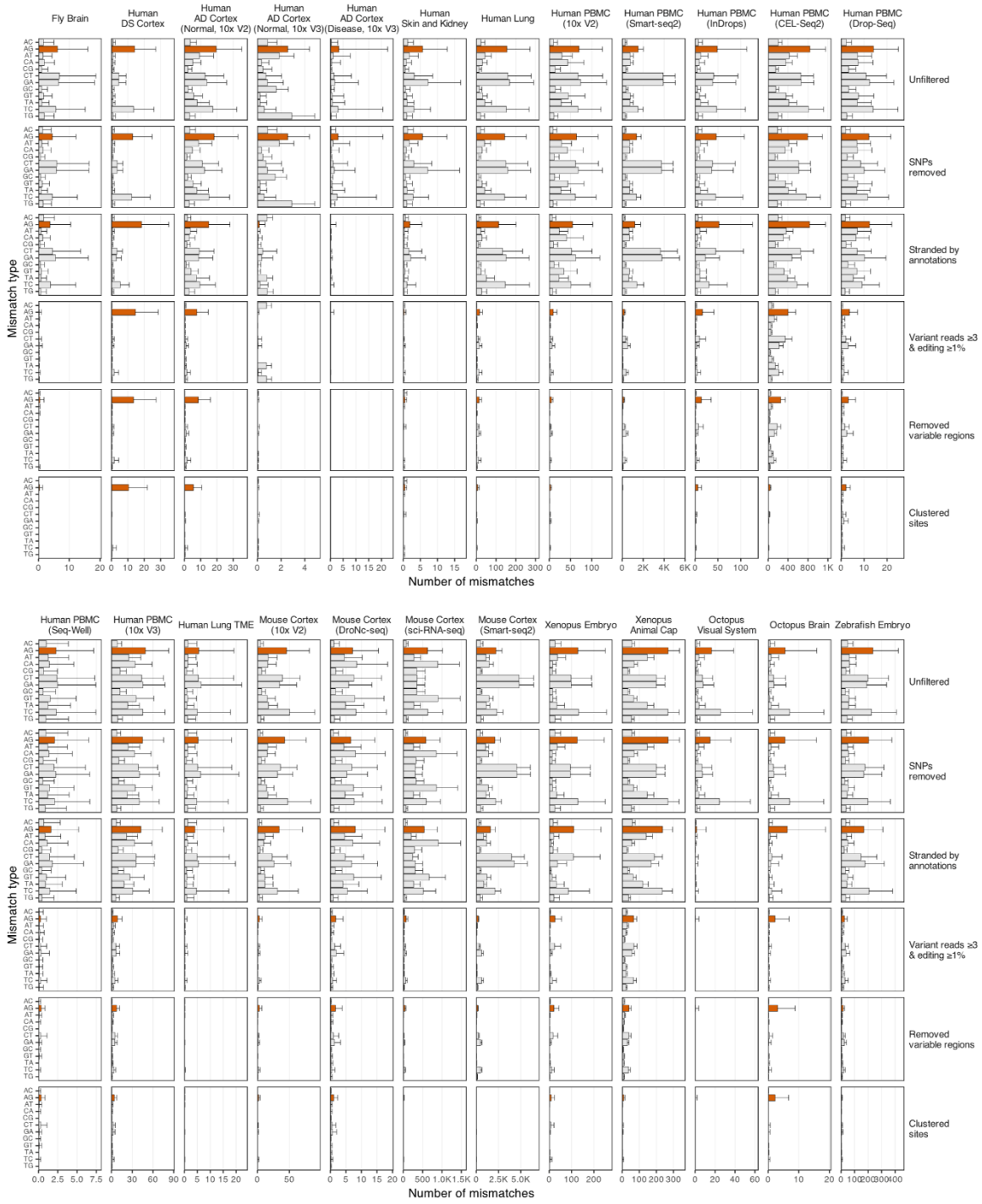

Figure S2| Site discovery at the single-cell level.
Bar plots showing the number of mismatches retained at each filtering step for high-confidence sites. Most panels are almost empty after filtering out low number of variant reads and editing rates, showing that single-cell level site discovery is infeasible due to insufficient coverage.

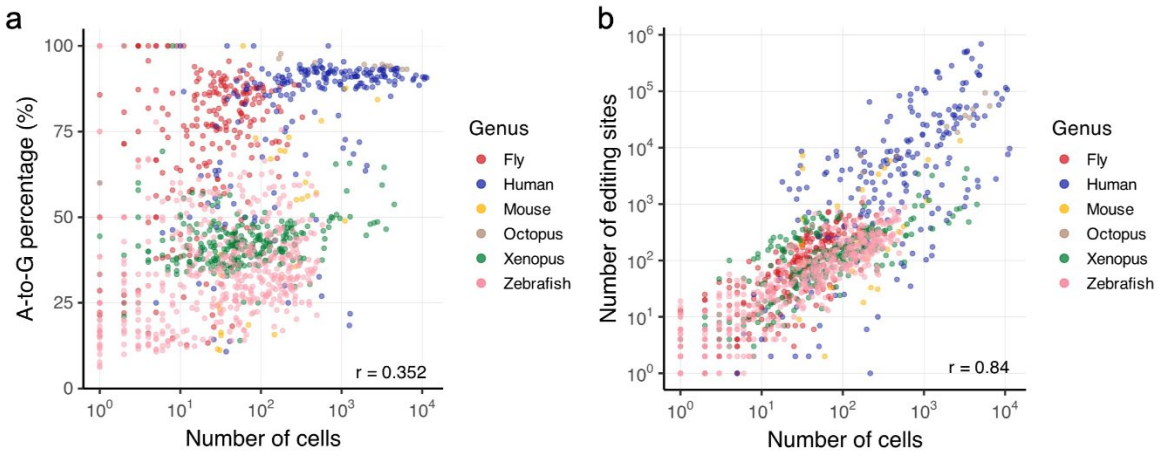

Figure S3| Effect of cell coverage on performance metrics.
Scatter plots showing the number of cells versus **(a)** A-to-G percentages and **(b)** number of sites. Correlation is indicated by the Pearson's r values on the plots.

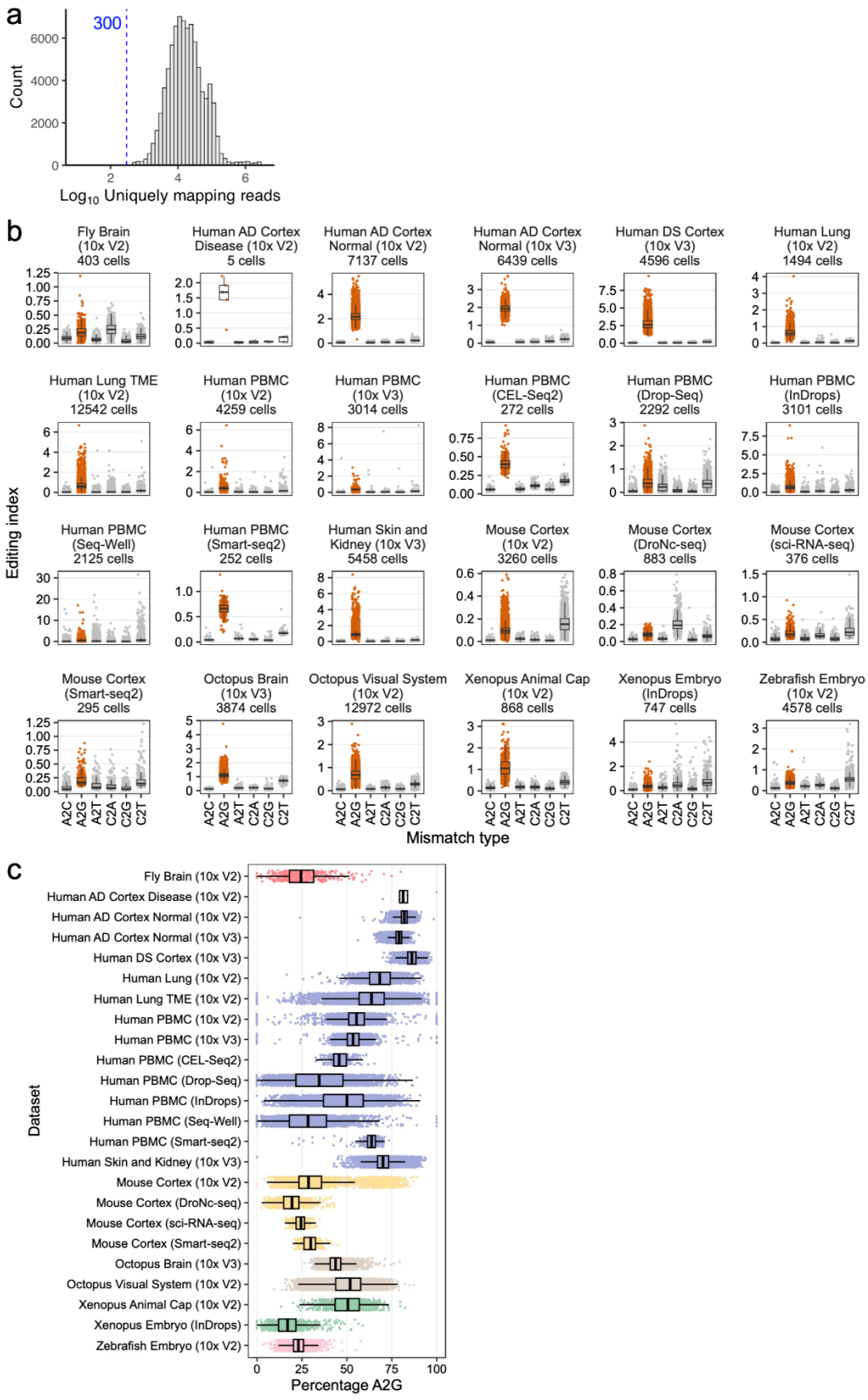

Figure S4| Quantification of per cell repeat editing indices for all datasets.
(a) Threshold for filtering out cells with low reads per cell. Blue dotted line indicates 300 reads, below which cells were dropped from downstream analyses. **(b)** Editing indices for all mismatch types per cell per dataset. **(c)** A-to-G percentages for all datasets coloured by genus.

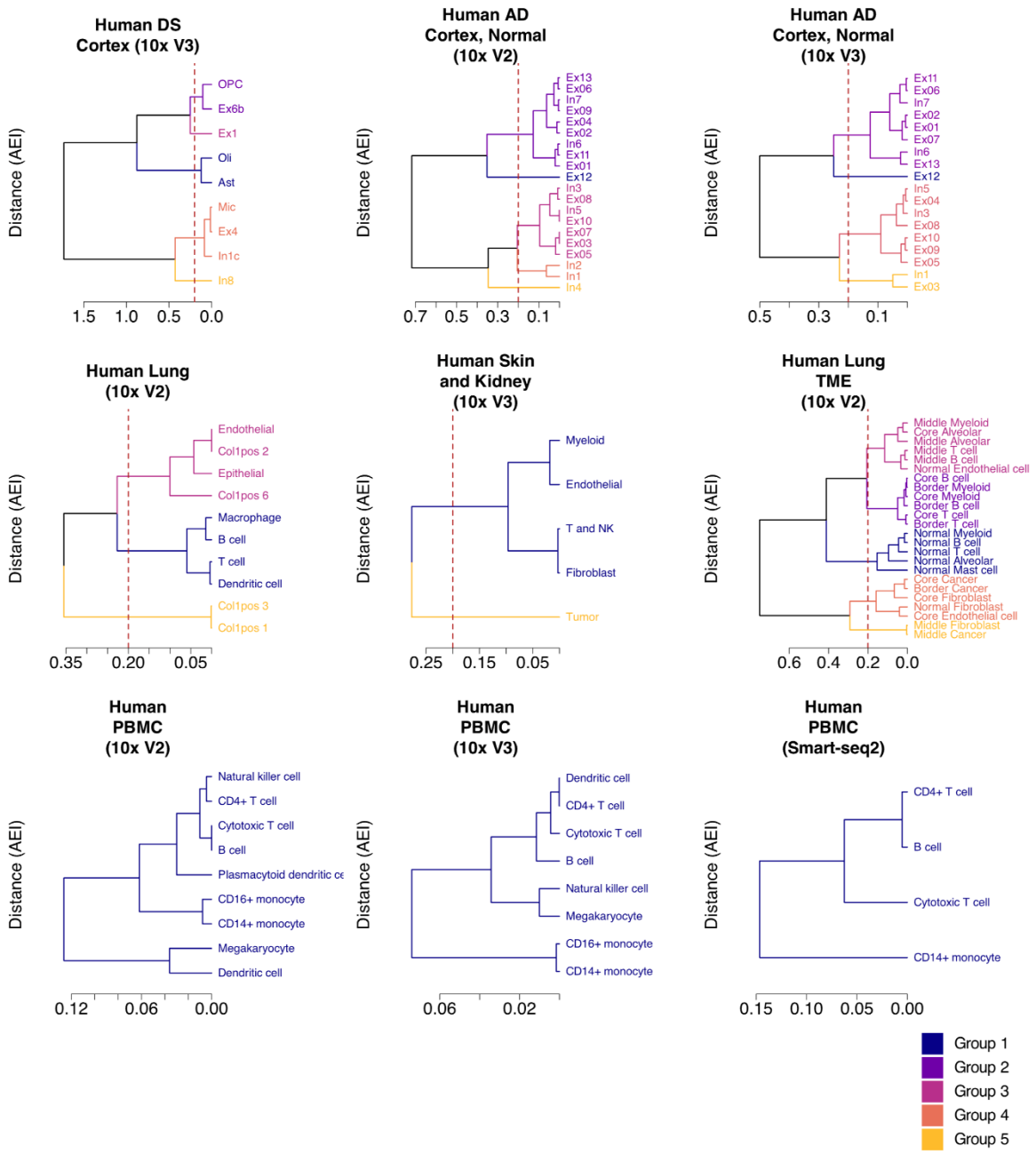

Figure S5| Hierarchical clustering results.
Hierarchical clustering of human 10x datasets showing that cell types can be grouped by AEI. Vertical red dotted line marks AEI = 0.2, where the dendrograms are cut to yield 1 to 5 groups.

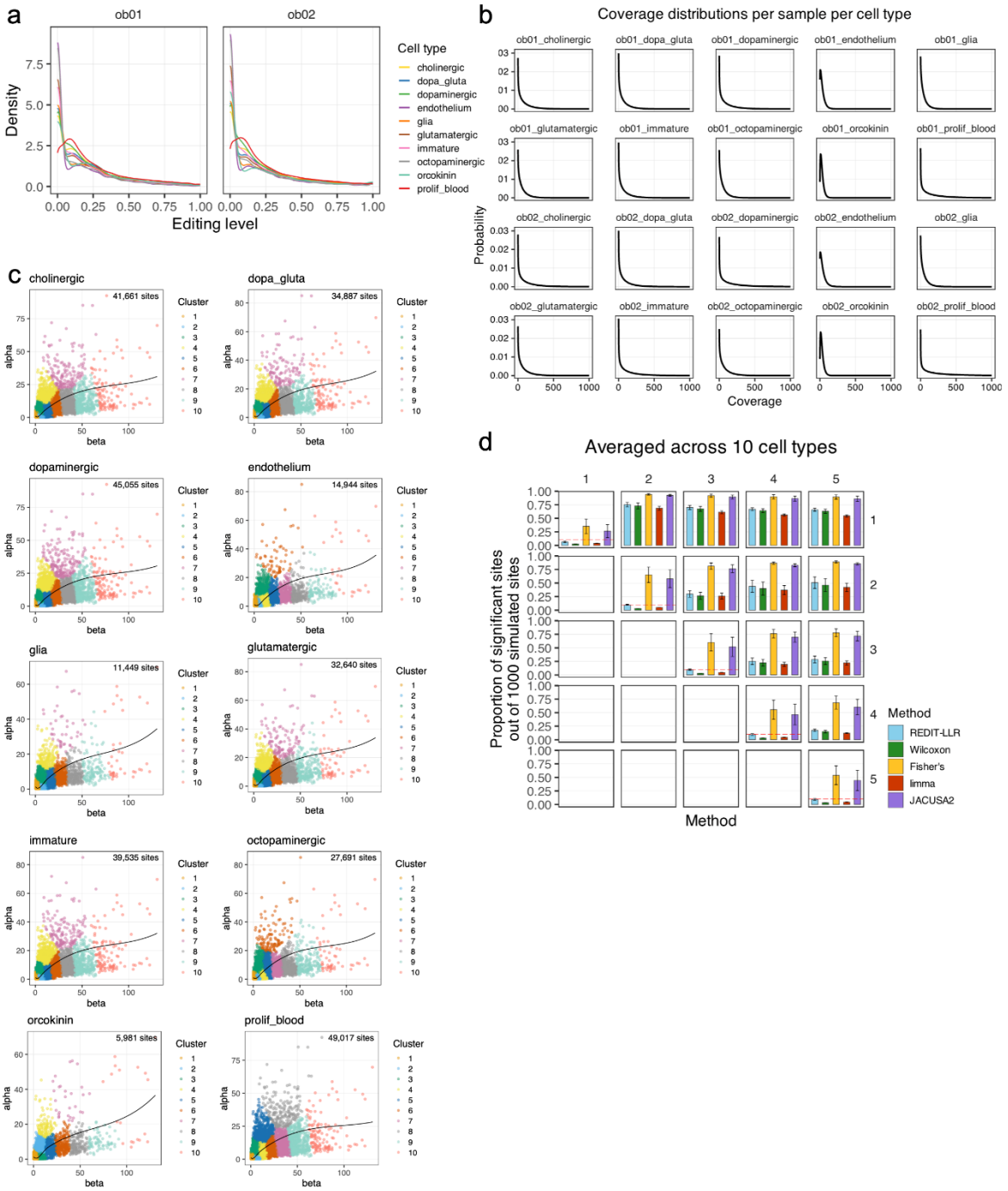

Figure S6| Additional results for differential editing analysis using simulated data. **(a)** Editing level distributions per cell type for each sample of the original *Octopus vulgaris* dataset. **(b)** Coverage distributions per sample per cell type from which coverage values were drawn for the simulations. **(c)** Editing levels in each cell type were split into 10 distribution clusters with k-means clustering. Cluster colours do not relate across cell types. **(d)** Simulation results for k = 5 instead of k = 10 when clustering editing distributions. Bars show mean ± SD.

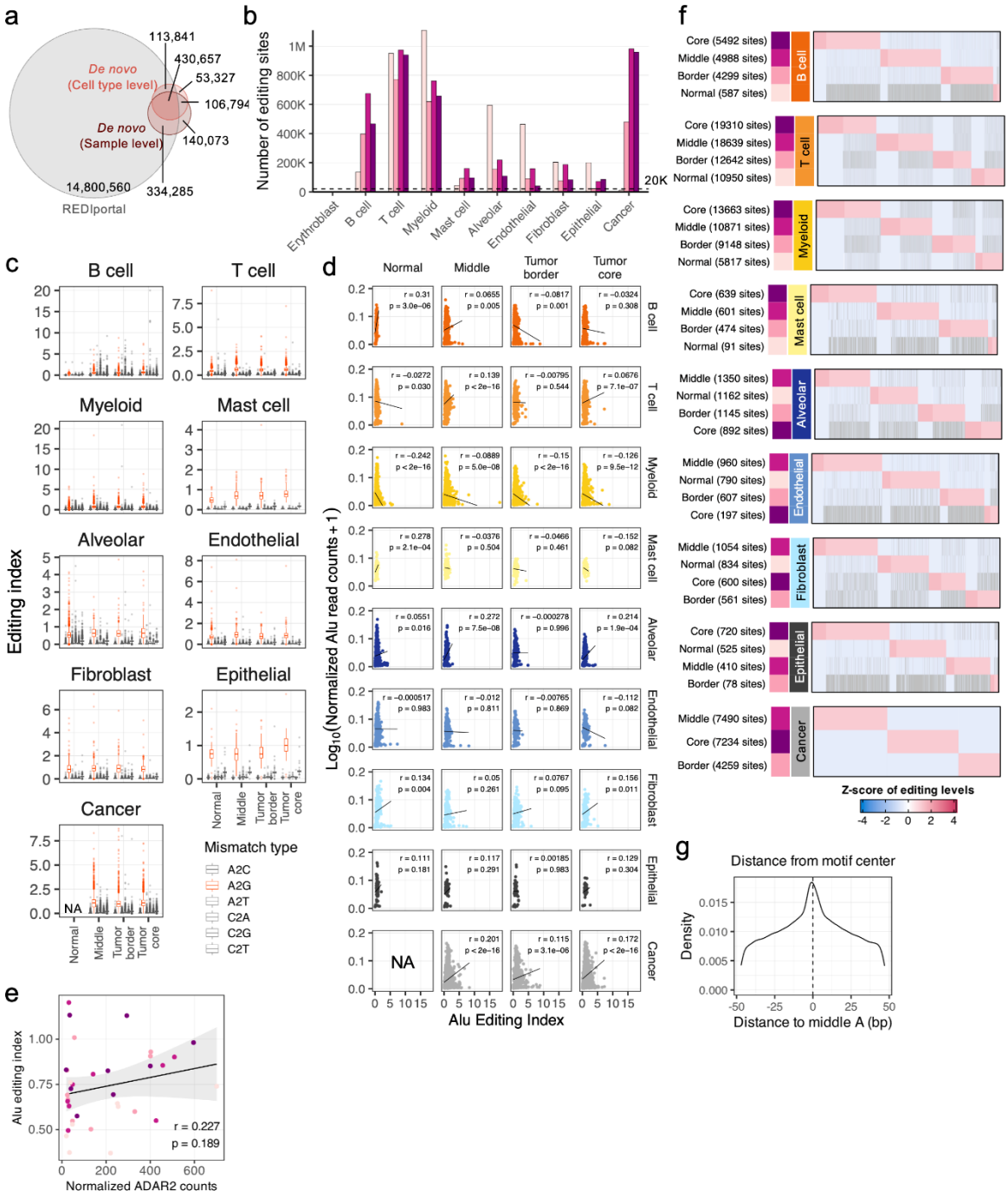
a
113,841
430,657
De novo
(Cell type level)
53,327
106,794
De novo
(Sample level)
140,073
14,800,560
334,285
REDIportal
b
Number of editing sites
1M
800K
600K
400K
200K
0
20K
Erythroblast
B cell
T cell
Myeloid
Mast cell
Alveolar
Endothelial
Fibroblast
Epithelial
Cancer
c
B cell
T cell
Myeloid
Mast cell
Alveolar
Endothelial
Fibroblast
Epithelial
Cancer
Editing index
NA
Normal
Middle
Tumor border
Tumor core
Mismatch type
A2C
A2G
A2T
C2A
C2G
C2T
d
Normal
Middle
Tumor border
Tumor core
B cell
T cell
Myeloid
Mast cell
Alveolar
Endothelial
Fibroblast
Epithelial
Cancer
Log10(Normalized Alu read counts + 1)
Alu Editing Index
e
Alu editing index
r = 0.227
p = 0.189
Normalized ADAR2 counts
f
Core (5492 sites)
Middle (4988 sites)
Border (4299 sites)
Normal (587 sites)
B cell
Core (19310 sites)
Middle (18639 sites)
Border (12642 sites)
Normal (10950 sites)
T cell
Core (13663 sites)
Middle (10871 sites)
Border (9148 sites)
Normal (5817 sites)
Myeloid
Core (639 sites)
Middle (601 sites)
Border (474 sites)
Normal (91 sites)
Mast cell
Middle (1350 sites)
Normal (1162 sites)
Border (1145 sites)
Core (892 sites)
Alveolar
Middle (960 sites)
Normal (790 sites)
Border (607 sites)
Core (197 sites)
Endothelial
Middle (1054 sites)
Normal (834 sites)
Core (600 sites)
Border (561 sites)
Fibroblast
Core (720 sites)
Normal (525 sites)
Middle (410 sites)
Border (78 sites)
Epithelial
Middle (7490 sites)
Core (7234 sites)
Border (4259 sites)
Cancer
Z-score of editing levels
g
Distance from motif center
Density
Distance to middle A (bp)

Figure S7| Additional results from our lung TME analysis.
(a) *De novo* site discovery identifies both known (REDIportal) and novel sites at sample and cell type-level. **(b)** Number of editing sites with ≥10 read coverage per cell type per region. Black dotted line marks 20,000 sites, below which the region-cell type combination was dropped from further analyses. **(c)** Global editing indices with A2G indices coloured in red, showing predominance of A2G mismatches across majority of cell types, indicating that AEI can be reliably quantified at single-cell level. **(d)** Increased AEI in each cell type was not due to increased *Alu* expression. Pearson's correlation coefficients (r) and p values are indicated in the respective panels. **(e)** Correlations of AEI with *ADAR2* levels, showing poor correlation. Pearson's correlation coefficient (r) and p value are indicated in the plot. **(f)** Heatmaps showing region-specific editing sites per cell type. Regions with highest number of specific sites are ordered first per cell type. Immune cell types show an ordered Core to Normal trend whereas non-immune cell types do not. **(g)** Distribution of RBP motifs around the background set of sites used for motif enrichment analysis.

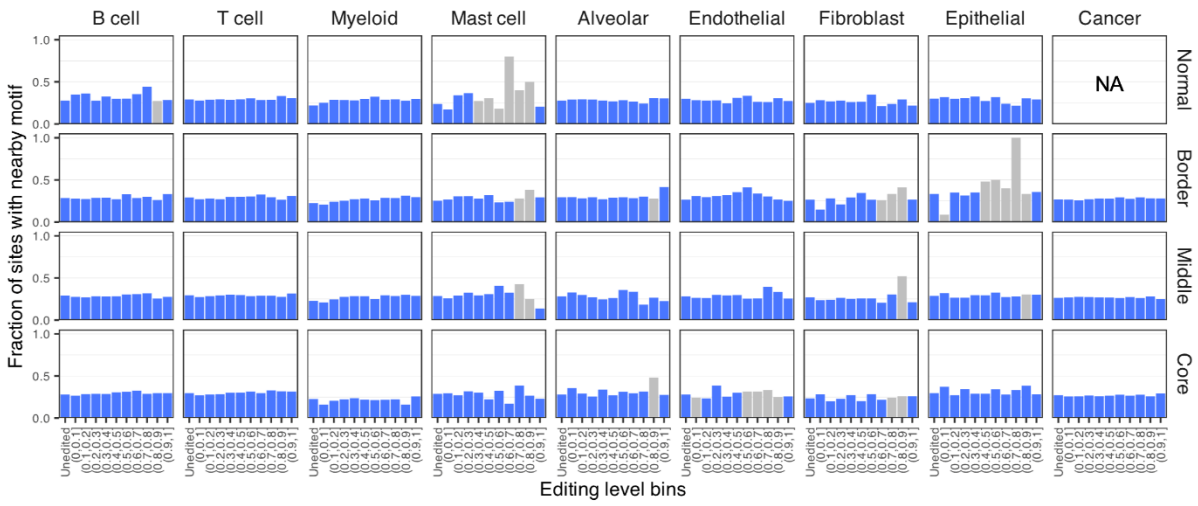

Figure S8| Motif presence near editing sites is independent of editing level.
Bar plots of the fraction of editing sites with at least one proximal motif, stratified by deciles of editing levels. Bars with fewer than 30 sites are shaded grey as low counts may exaggerate observed fractions.

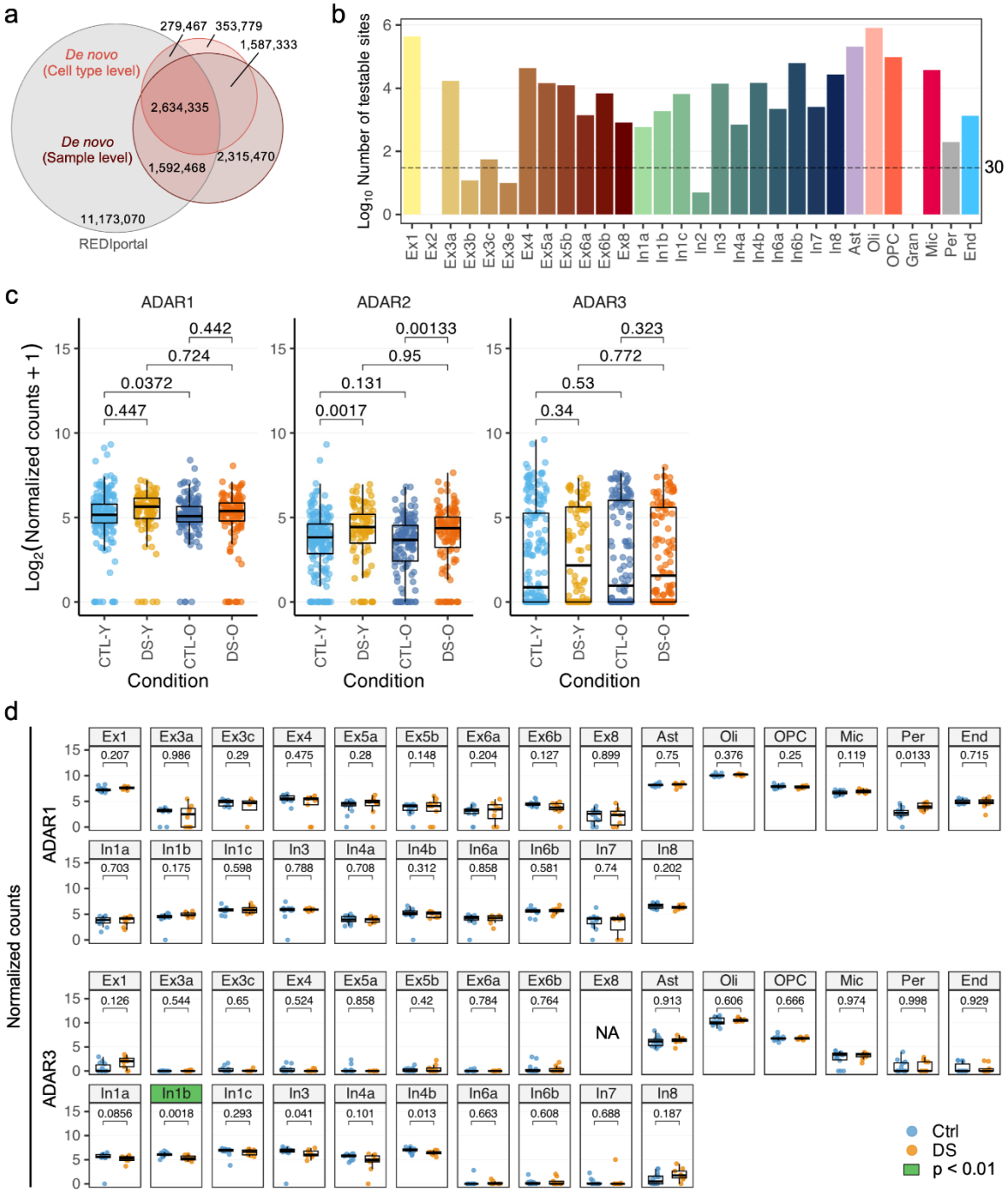

Figure S9| Site discovery and *ADAR* expression in the primary Down syndrome dataset. **(a)** Overlaps of *de novo* discovered sites with REDIportal sites for both sample and cell type level discovery strategies. **(b)** Bar plots showing the number of editing sites that passed minimum thresholds for differential testing when stratified by disease state. Black dotted line indicates 30 sites, below which a cell type was excluded from further analyses. **(c)** *ADAR1*, *ADAR2*, and *ADAR3* transcript levels per cell type among different age and disease state groups. CTL-Y: Control Young; DS-Y: DS Young; CTL-O: Control Old; DS-O: DS Old. DESeq2 Wald test p values are indicated for each respective comparison. **(d)** Box plots comparing *ADAR1* and *ADAR3* levels between Ctrl and DS per cell type. Green cell type name labels indicate significance (DESeq2 Wald test, $p < 0.01$). NA indicates no detectable reads for all samples of that cell type.

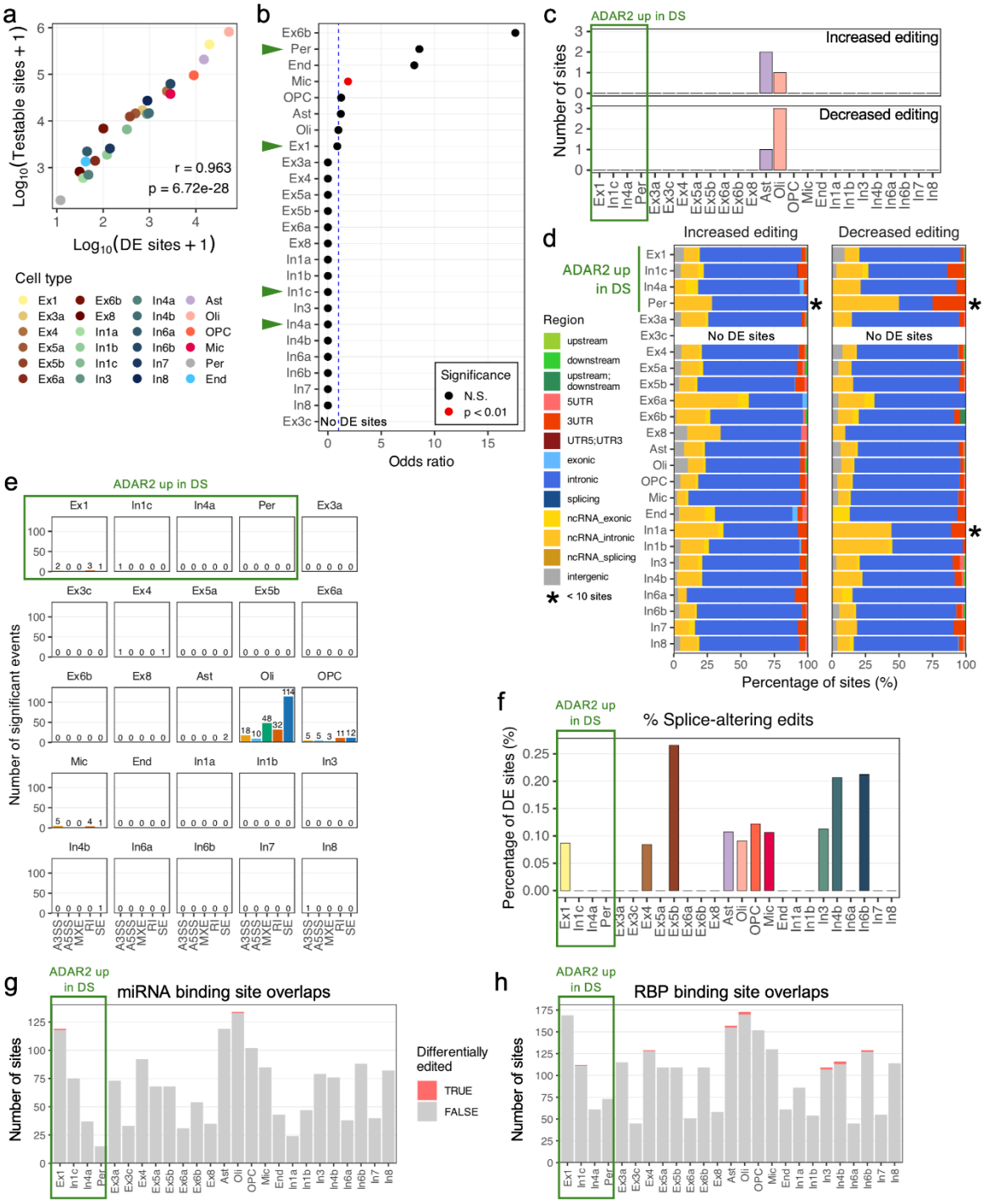
a
Log10(Testable sites + 1)
Log10(DE sites + 1)
r = 0.963
p = 6.72e-28
Cell type
Ex1 Ex3a Ex4 Ex5a Ex5b Ex6a Ex6b Ex8 In1a In1b In1c In3 In4a In4b In6a In6b In7 In8 Ast Oli OPC Mic Per End
b
Significance
N.S.
p < 0.01
No DE sites
Odds ratio
c
ADAR2 up in DS
Increased editing
Decreased editing
Number of sites
d
ADAR2 up in DS
Increased editing
Decreased editing
Region
upstream
downstream
upstream; downstream
5UTR
3UTR
UTR5;UTR3
exonic
intronic
splicing
ncRNA_exonic
ncRNA_intronic
ncRNA_splicing
intergenic
< 10 sites
No DE sites
Percentage of sites (%)
e
ADAR2 up in DS
Number of significant events
A3SS A5SS MXE RI SE
f
ADAR2 up in DS
% Splice-altering edits
Percentage of DE sites (%)
g
ADAR2 up in DS
miRNA binding site overlaps
Number of sites
Differentially edited
TRUE
FALSE
h
ADAR2 up in DS
RBP binding site overlaps
Number of sites

Figure S10| Effects of differentially edited sites from pooled analyses of the primary Down syndrome dataset.
(a) Log-transformed counts of differentially edited sites and initial number of testable sites per cell type. Pearson's correlation coefficient (r) and p value are indicated on the plot. **(b)** Odds ratios from association tests for enrichment of differentially expressed and edited genes per cell type. Red dots indicate significance from Fisher's exact tests ($p < 0.01$). **(c)** Number of the known 1,517 CDS sites that were differentially edited per cell type. **(d)** Proportion of differentially edited sites within different gene regions. Asterisks indicate cell types with fewer than 10 differentially edited sites. **(e)** Number of significant splicing events per cell type from rMATS-turbo. A3SS: alternative 3' splice site; A5SS: alternative 5' splice site; MXE: mutually exclusive exons; RI: retained intron; SE: Skipped exon. **(f)** Percentage of differentially edited sites that have potential splice-altering effects (SpliceAI score ≥ 0.2). Number of editing sites overlapping **(g)** microRNA and **(h)** RBP binding sites. Green arrows, boxes, and markups indicate cell types with increased *ADAR2* expression as in **Figure 7b** for panels **(b-h)**.

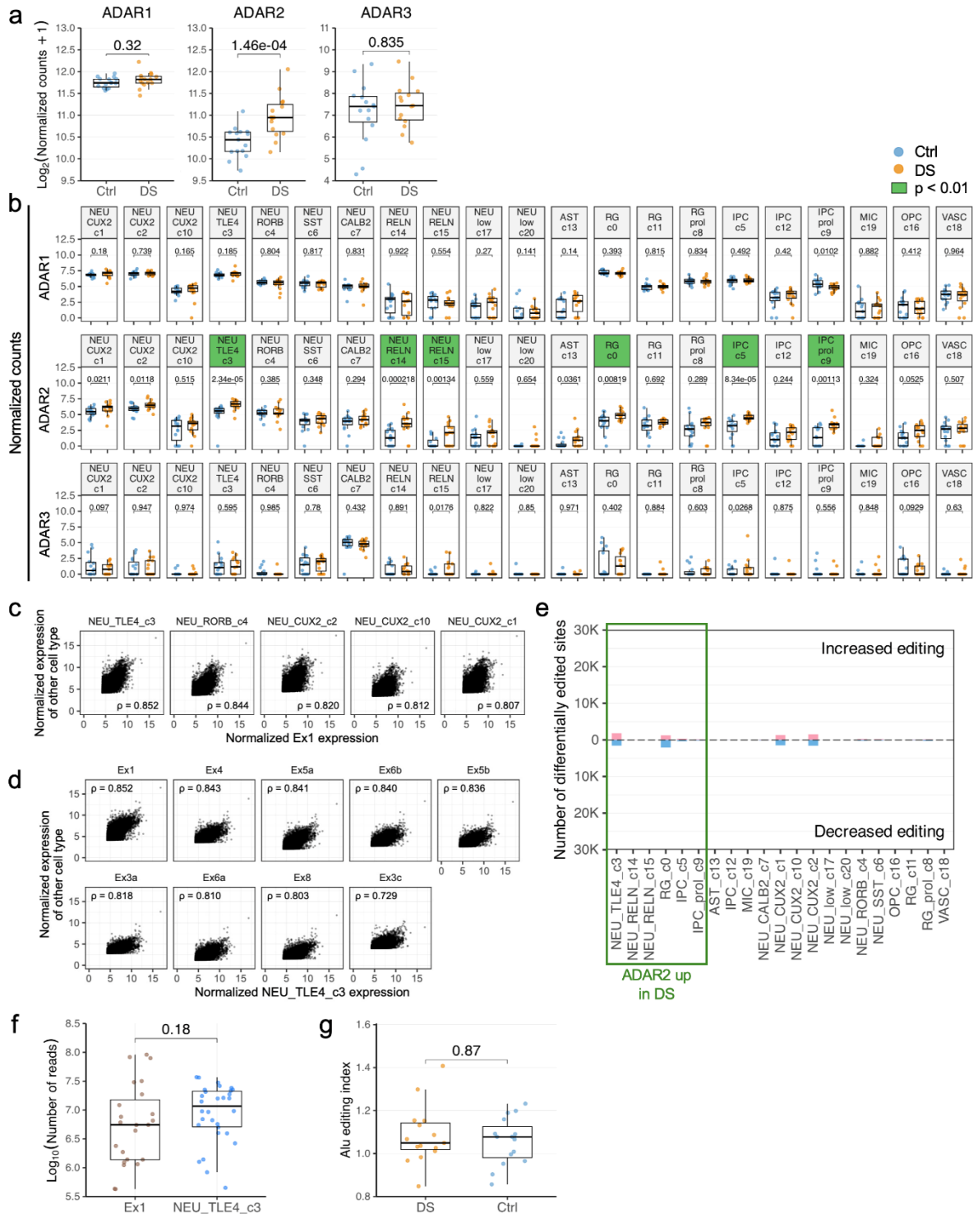
a
ADAR1
ADAR2
ADAR3
0.32
1.46e-04
0.835
Log2(Normalized counts + 1)
Ctrl
DS
b
Normalized counts
ADAR1
ADAR2
ADAR3
Ctrl
DS
p < 0.01
c
Normalized expression of other cell type
Normalized Ex1 expression
NEU_TLE4_c3
NEU_RORB_c4
NEU_CUX2_c2
NEU_CUX2_c10
NEU_CUX2_c1
ρ = 0.852
ρ = 0.844
ρ = 0.820
ρ = 0.812
ρ = 0.807
d
Normalized NEU_TLE4_c3 expression
Ex1
Ex4
Ex5a
Ex6b
Ex5b
Ex3a
Ex6a
Ex8
Ex3c
ρ = 0.852
ρ = 0.843
ρ = 0.841
ρ = 0.840
ρ = 0.836
ρ = 0.818
ρ = 0.810
ρ = 0.803
ρ = 0.729
e
Number of differentially edited sites
Increased editing
Decreased editing
ADAR2 up in DS
f
Log10(Number of reads)
0.18
Ex1
NEU_TLE4_c3
g
Alu editing index
0.87
DS
Ctrl

Figure S11| Analysis results for the fetal Down Syndrome dataset.
(a) *ADAR1*, *ADAR2*, and *ADAR3* expression levels between control (Ctrl) and DS samples. DESeq2 Wald test p values are indicated. **(b)** Per cell type comparisons of *ADAR1*, *ADAR2*, and *ADAR3* levels between Ctrl and DS samples. Green cell type name labels indicate significance (DESeq2 Wald test, $p < 0.01$). **(c)** Correlations of transcriptional profiles between Ex1 neurons from the primary dataset and all excitatory neuron subtypes of the fetal dataset. **(d)** Correlations of transcriptional profiles between NEU_TLE4_c3 neurons from the fetal DS dataset and all excitatory neuron subtypes of the primary DS dataset. Spearman's ρ for each correlation are indicated on the respective plots in **(c-d)**. **(e)** Number of differentially edited sites between Ctrl and DS per cell type. Green box marks out cell types with increased *ADAR2* levels as in **(b)**. **(f)** Sequencing depths for NEU_TLE_c3 neurons from the fetal dataset and Ex1 neurons from the primary dataset. **(g)** AEI levels between Ctrl and DS samples.

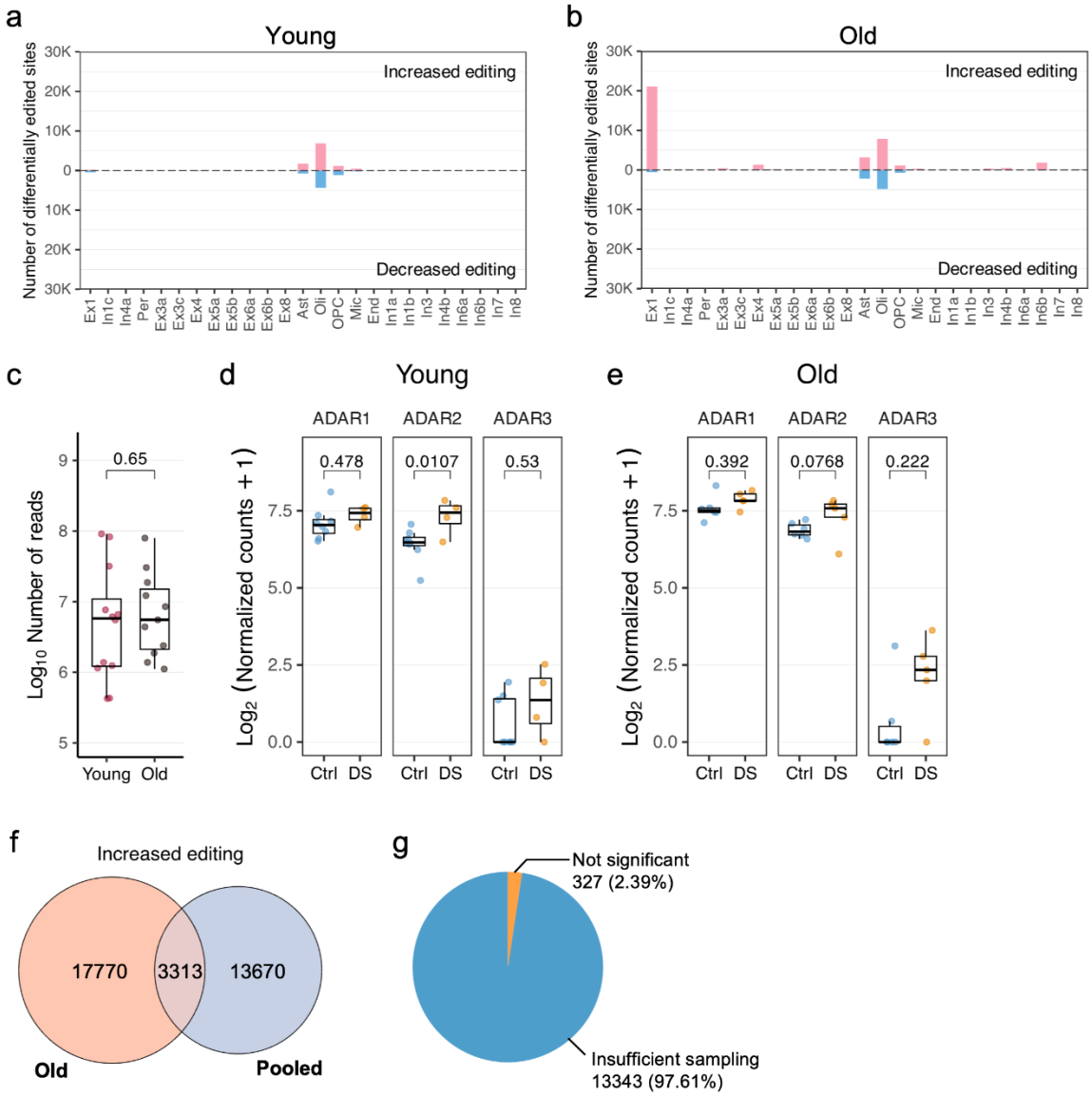

Figure S12| Additional results from analyses of the primary Down syndrome dataset. **(a-b)** Number of differentially edited sites per cell type in **(a)** young and **(b)** old samples from the primary dataset. **(c)** Number of reads in young and old samples for Ex1. Wilcoxon p value is indicated in the plot. **(d-e)** Comparisons of *ADAR1*, *ADAR2*, and *ADAR3* levels between Ctrl and DS Ex1 neurons in **(d)** young and **(e)** old samples. DESeq2 Wald test p values are indicated. **(f)** Overlaps between the number of differentially up-regulated editing sites in Ex1 using old samples and pooled samples. **(g)** Reasons why the 13,670 pooled-analysis sites in **(f)** were not detected in old-only analysis.

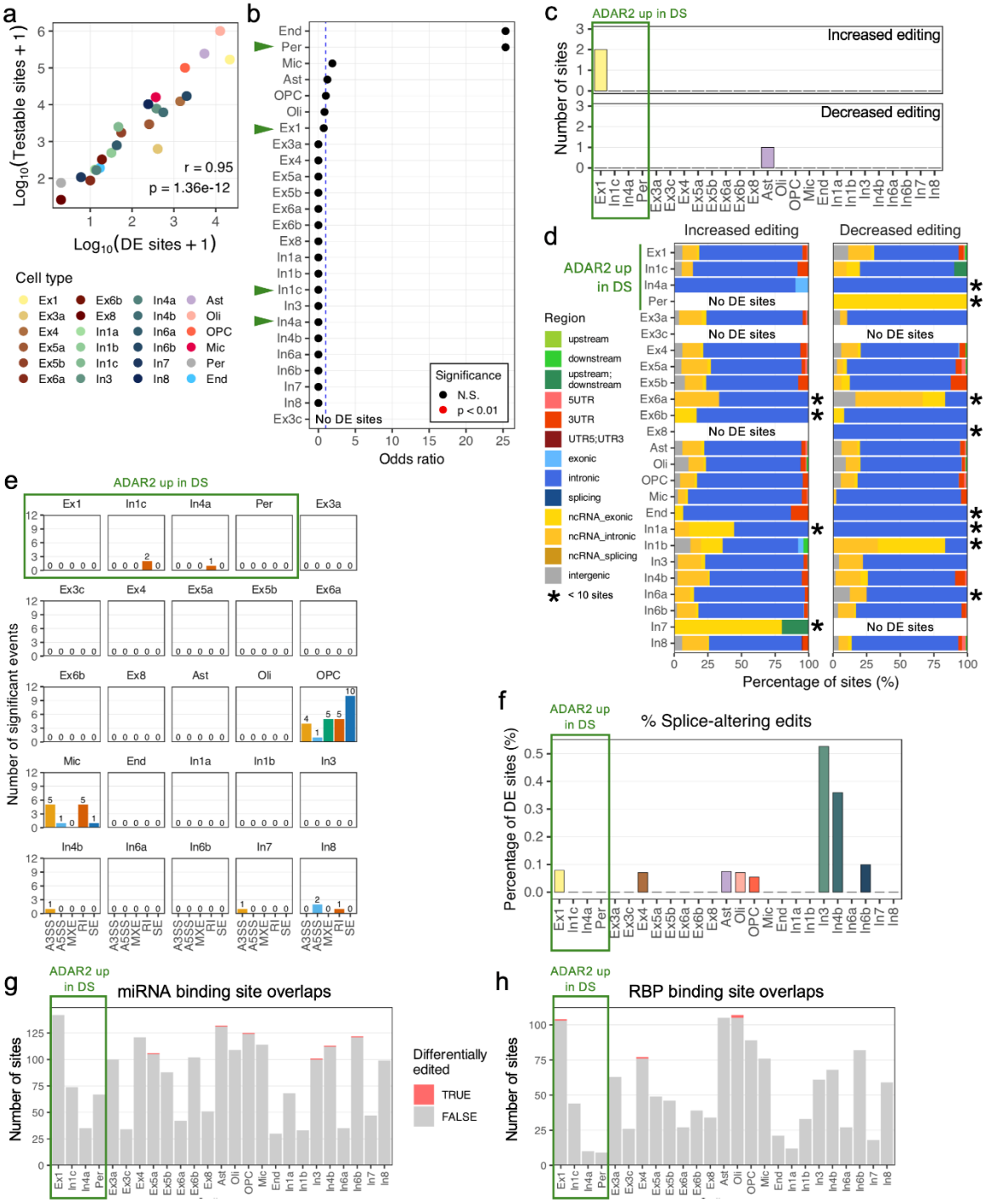
a
Log10(Testable sites + 1)
Log10(DE sites + 1)
r = 0.95
p = 1.36e-12
Cell type
Ex1 Ex3a Ex4 Ex5a Ex5b Ex6a Ex6b Ex8 In1a In1b In1c In3 In4a In4b In6a In6b In7 In8 Ast Oli OPC Mic Per End
b
No DE sites
Significance
N.S.
p < 0.01
Odds ratio
c
ADAR2 up in DS
Increased editing
Decreased editing
Number of sites
d
ADAR2 up in DS
Increased editing
Decreased editing
Region
upstream
downstream
upstream; downstream
5UTR
3UTR
UTR5;UTR3
exonic
intronic
splicing
ncRNA_exonic
ncRNA_intronic
ncRNA_splicing
intergenic
< 10 sites
No DE sites
Percentage of sites (%)
e
ADAR2 up in DS
Number of significant events
A3SS A5SS MXE RI SE
f
ADAR2 up in DS
% Splice-altering edits
Percentage of DE sites (%)
g
ADAR2 up in DS
miRNA binding site overlaps
Number of sites
Differentially edited
TRUE
FALSE
h
ADAR2 up in DS
RBP binding site overlaps
Number of sites

Figure S13| Analysis of differentially edited sites from old samples alone.
(a) Log-transformed counts of differentially edited sites and initial number of testable sites per cell type. Pearson's correlation coefficient (r) and p value are indicated on the plot. **(b)** Odds ratios from association tests for enrichment of differentially expressed and edited genes per cell type. Red dots indicate significance from Fisher's exact tests ($p < 0.01$). **(c)** Number of the known 1,517 CDS sites that were differentially edited per cell type. **(d)** Proportion of differentially edited sites within different gene regions. Asterisks indicate cell types with fewer than 10 differentially edited sites. **(e)** Number of significant splicing events per cell type from rMATS-turbo. A3SS: alternative 3' splice site; A5SS: alternative 5' splice site; MXE: mutually exclusive exons; RI: retained intron; SE: Skipped exon. **(f)** Percentage of differentially edited sites that have potential splice-altering effects (SpliceAI score ≥ 0.2). Number of editing sites overlapping **(g)** microRNA and **(h)** RBP binding sites. Green arrows, boxes, and markups indicate cell types with increased *ADAR2* expression as in **Figure 7b** for panels **(b-h)**.

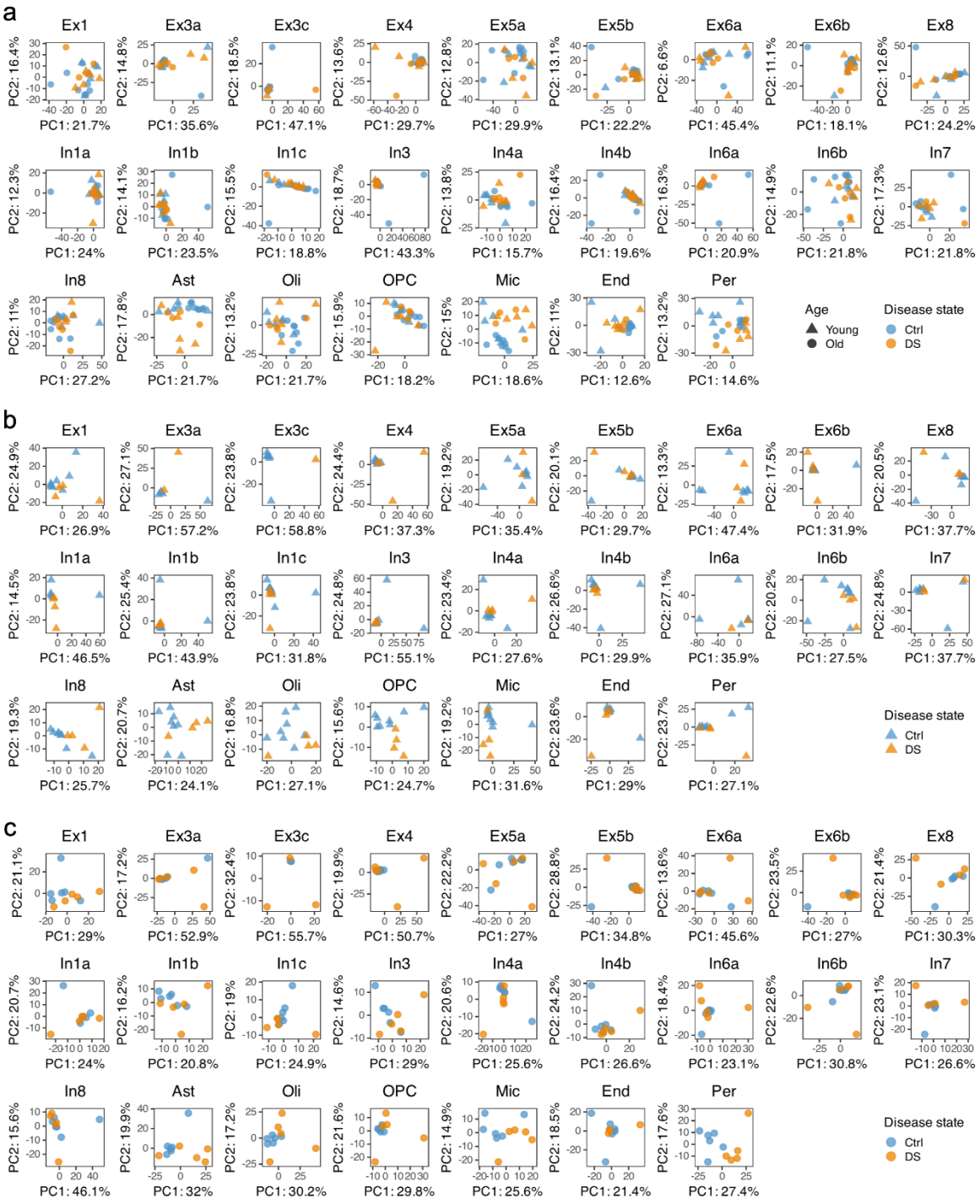

Figure S14| Principal component analysis of gene expression.
Principal component analysis (PCA) plots per cell type showing the top two PCs from analysing **(a)** pooled, **(b)** young, and **(c)** old samples. For majority of cell types, Ctrl and DS samples do not segregate clearly along PC1 or PC2.

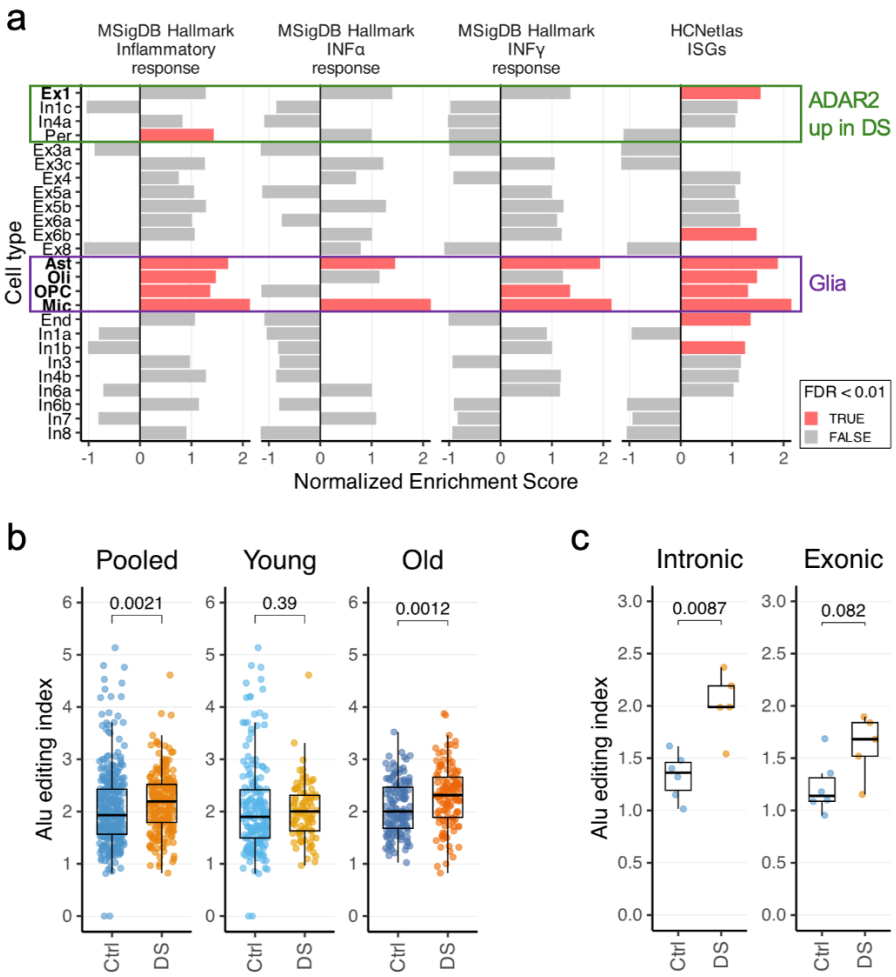

Figure S15| GSEA results for young samples and AEI differences.
(a) Normalized enrichment scores from gene set analysis results per cell type for young samples. Red colour indicates significance (FDR < 0.01). Green boxes mark out cell types with increased *ADAR2* in DS as in **Figure 7b**, while purple boxes mark out the glia cell types. **(b)** Box plots comparing overall AEI levels per sample per cell type, for pooled, young only, and old only analyses. **(c)** Box plots showing per sample AEI quantified over all (left) intronic and (right) exonic *Alu* repeats in old pericytes.