Supplementary Figures:

Genome-wide Identification and Characterisation of Tissue-specific RNA Editing Events in *D. melanogaster* and their Potential Role in Regulating Alternative Splicing

Alborz Mazloomian and Irmtraud M. Meyer

Data Set	Tissue	Data Set	Tisue	Data Set	Tissue
Mod4241	Head	Mod4266	Ovaries	Mod4259	Digestive System
Mod4242	Head	Mod4247	Accessory Glands	Mod4256	Central Nervous System
Mod4243	Head	Mod4249	Testes	Mod4257	Central Nervous System
Mod4245	Head	Mod4250	Carcass	Mod4260	Fat Body
Mod4246	Head	Mod4252	Carcass	Mod4267	Fat Body
Mod4248	Head	Mod4254	Carcass	Mod4268	Fat Body
Mod4263	Head	Mod4258	Carcass	Mod4261	Imaginal Discs
Mod4264	Head	Mod4251	Digestive System	Mod4262	Salivary Glands
Mod4265	Head	Mod4253	Digestive System	Mod4269	Salivary Glands
Mod4244	Ovaries	Mod4255	Digestive System		

Figure S1. Tissue specific data sets selected from the modENCODE project. The IDs of the selected libraries from the modENCODE project and the tissues from which these libraries are sampled from, are shown here. The data contain 29 libraries from 10 tissues. Reads in each library are from different developmental stages.

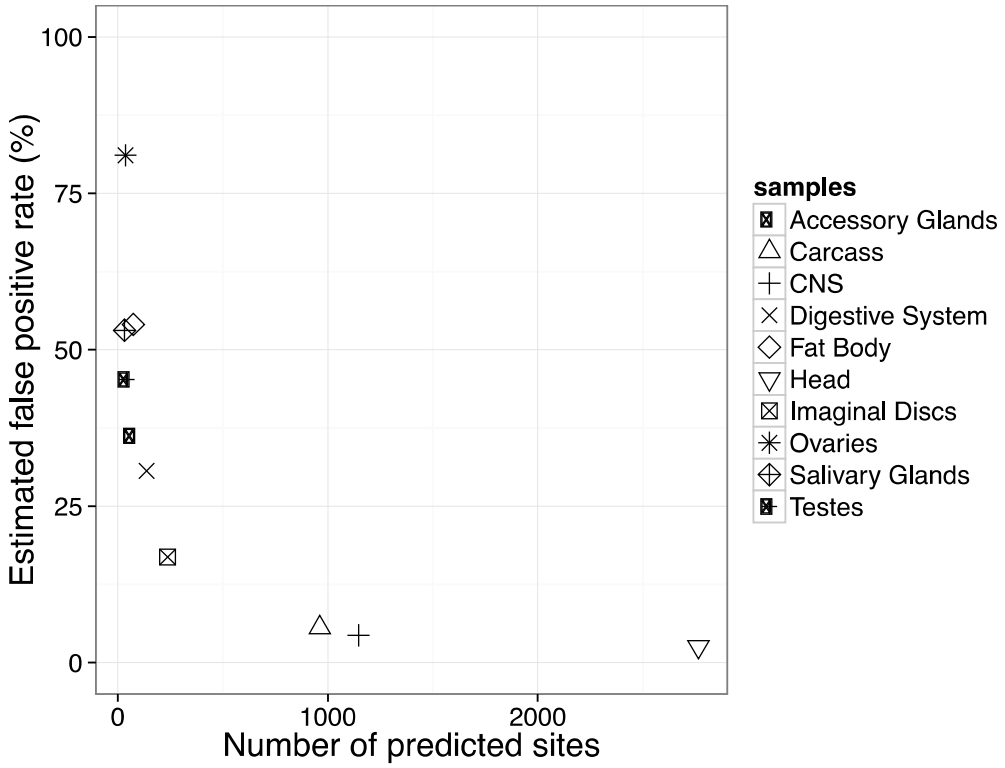

Figure S2.

Estimated error rate versus the number of predicted sites in different tissues of our study. To roughly evaluate the performance of our pipeline, we counted the number of A-to-G and the number of G-to-A conversions for each tissue. Assuming that all G-to-A events are errors or heterozygous sites, and assuming that the same number of A-to-G errors or heterozygous sites, we estimated the expected error rate of our predictions for each tissue. Y axis shows the estimated error and X axis the number of detected sites per tissue.

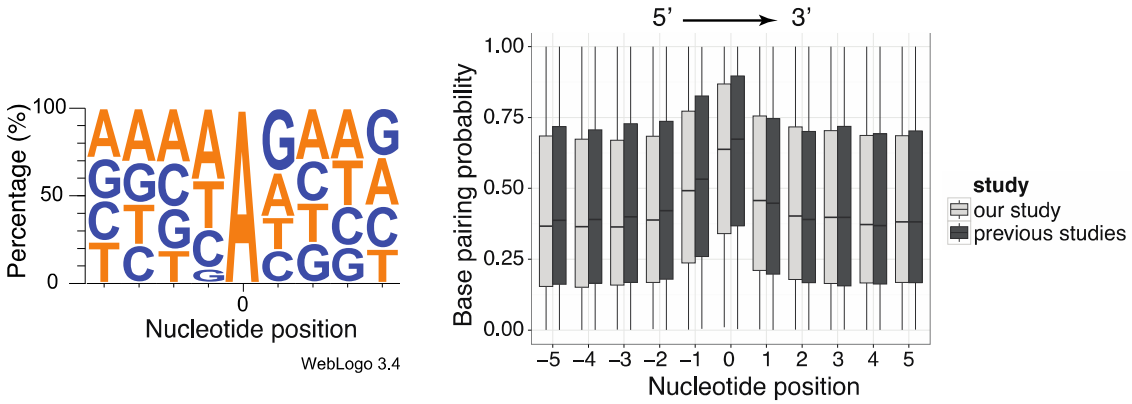

Figure S3. ADAR sequence and structural preferences. **(A)** The frequency of each nucleotide at each position relative to the predicted editing sites. Guanosine is depleted at the exact 5' position of editing sites. **(B)** Average base pairing probabilities computed using RNAplfold[1] for regions close to ADAR targets for sites predicted in our study, and previous studies.[2-5] Positions -1 to 1 show higher average pairing probabilities compared to other loci. Using structural features in our pipeline may bias our predictions towards sites with higher base pairing probabilities around reported sites; however a similar pattern has also been observed when considering sites predicted in previous studies. Part **(A)** is generated using WEBLOGO.3.4 [http://weblogo.berkeley.edu/logo.cgi]

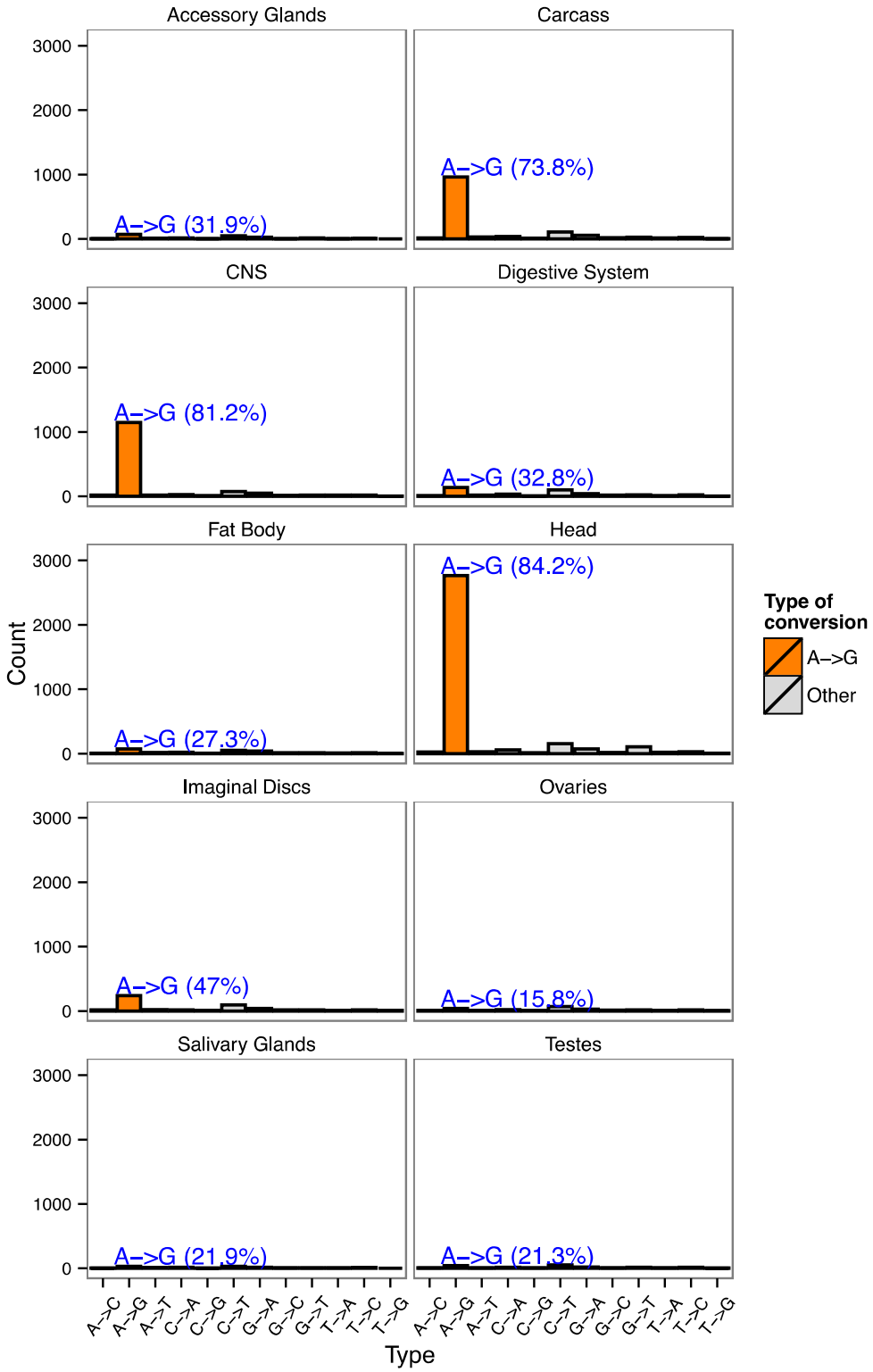
Accessory Glands
Carcass
CNS
Digestive System
Fat Body
Head
Imaginal Discs
Ovaries
Salivary Glands
Testes
A–>G (31.9%)
A–>G (73.8%)
A–>G (81.2%)
A–>G (32.8%)
A–>G (27.3%)
A–>G (84.2%)
A–>G (47%)
A–>G (15.8%)
A–>G (21.9%)
A–>G (21.3%)
Count
3000
2000
1000
0
A–>C
A–>G
A–>T
C–>A
C–>G
C–>T
G–>A
G–>C
G–>T
T–>A
T–>C
T–>G
Type
Type of conversion
A–>G
Other

Figure S4. Number of different conversion types in tissues of the modENCODE project. For most of the tissues, editing is rare, and most of our sites occur in head and CNS (central nervous system).

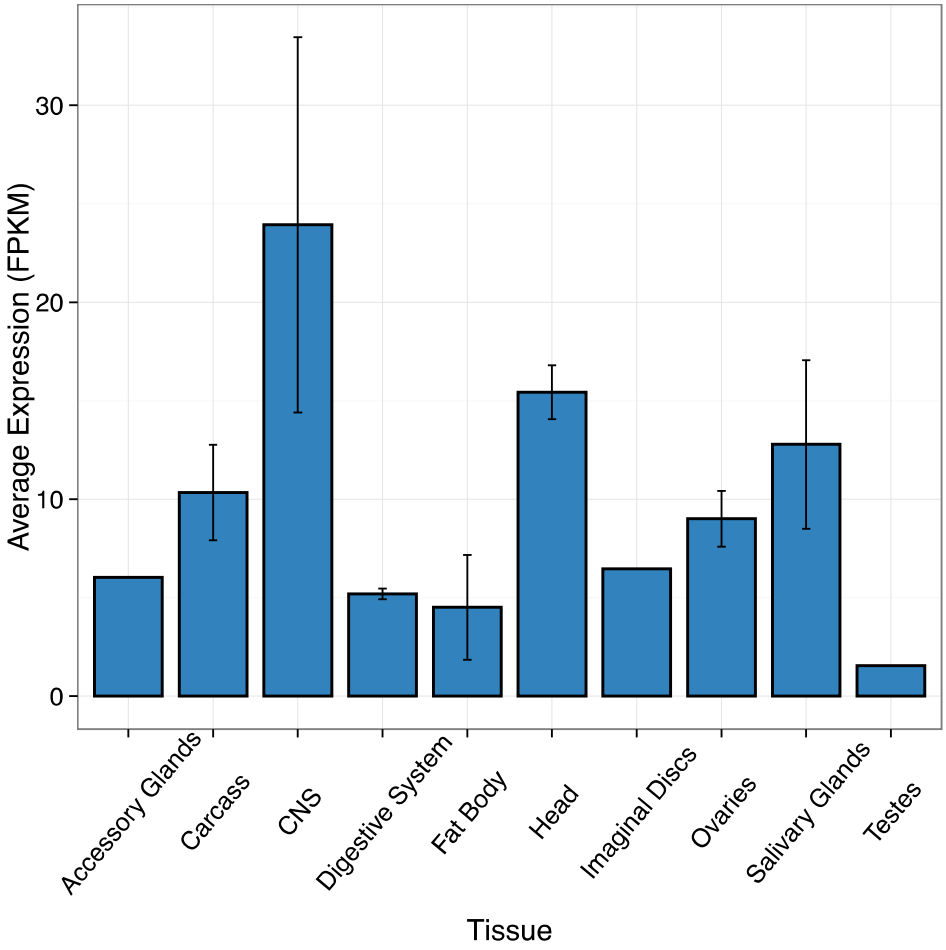

Figure S5. Average expression of dADAR in tissues of the modENCODE project. Expression values are measured in FPKM (fragments per kilobase of transcript per million fragments mapped) unit using CUFFLINKS[6] version "2.2.1". Although dADAR expression is highest in CNS (central nervous system) and head, but the gene is expressed in other tissues as well.

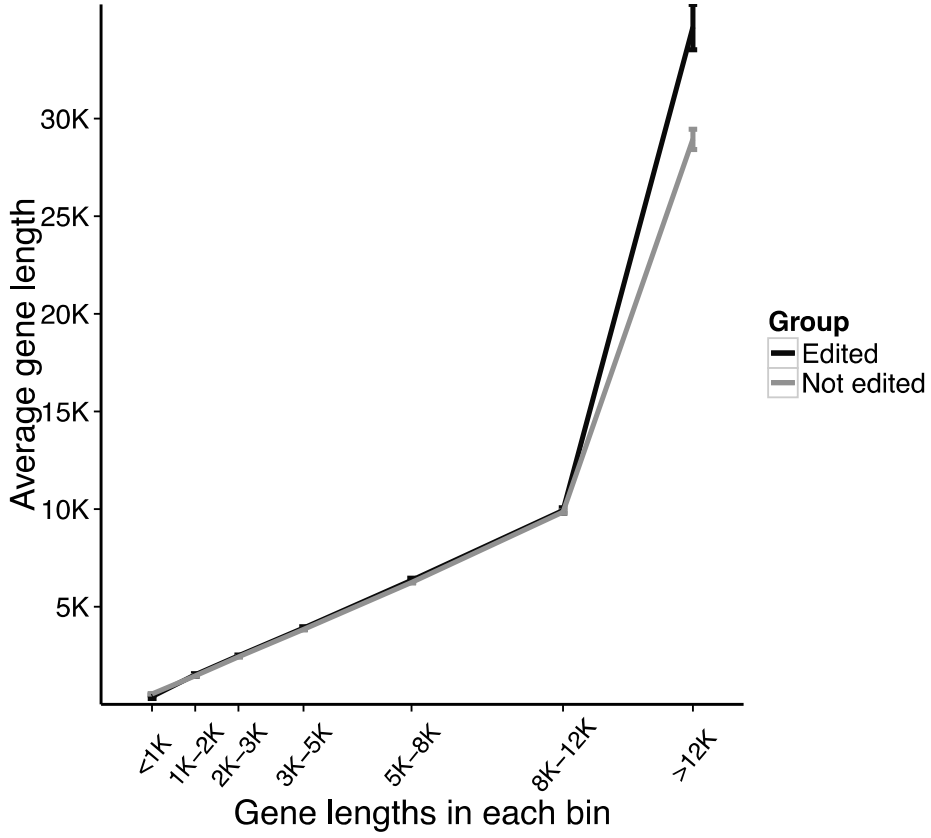

Figure S6. Average gene length for each bin in our gene set. To check if edited genes contain more annotated isoforms on average compared to un-edited genes, we split genes into bins based on their lengths (see main text). Here, we tested whether genes in the same bin from edited and un-edited group, have similar lengths. The plot shows that for most of our bins, average gene length is almost equal for edited and un-edited group.

References.

1. Bernhart SH, Hofacker IL, Stadler PF. Local RNA base pairing probabilities in large sequences. Bioinformatics 2006; 22:614-5.

2. Ramaswami G, Zhang R, Piskol R, Keegan LP, Deng P, O'Connell MA, Li JB. Identifying RNA editing sites using RNA sequencing data alone. Nature Methods 2013; 10:128-32.

3. Rodriguez J, Menet JS, Rosbash M. Nascent-seq indicates widespread cotranscriptional RNA editing in Drosophila. Molecular Cell 2012; 47:27-37.

4. St Laurent G, Tackett MR, Nechkin S, Shtokalo D, Antonets D, Savva YA, Maloney R, Kapranov P, Lawrence CE, Reenan RA. Genome-wide analysis of A-to-I RNA editing by single-molecule sequencing in Drosophila. Nature Structural & Molecular Biology 2013; 20:1333-9.

5. Graveley BR, Brooks AN, Carlson JW, Duff MO, Landolin JM, Yang L, Artieri CG, van Baren MJ, Boley N, Booth BW, et al. The developmental transcriptome of Drosophila melanogaster. Nature 2011; 471:473-9.

6. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nature Protocols 2012; 7:562-78.