SI Appendix

Alternative splicing events as peripheral biomarkers for motor learning deficit caused by adverse prenatal environments

Dipankar J. Dutta[1,8], Junko Sasaki[1,2,8], Ankush Bansal[1], Keiji Sugai[1,2], Satoshi Yamashita[1], Guojiao Li[2], Christopher Lazarski[3], Wang Li[1], Toru Sasaki[4], Chiho Yamashita[1], Heather Carryl[1], Ryo Suzuki[2], Masato Odawara[2], Yuka Imamura Kawasawa[5], Pasko Rakic[6,✉], Masaaki Torii[1,7,✉], Kazue Hashimoto-Torii[1,7, ✉]

The SI appendix includes:

- Supplementary Materials and Methods
- Supplementary Fig. S 1 to 20
- Supplementary Table S1
- Legend for Supplementary Movie S1
- Supplementary References

Other Supplemental Materials for this manuscript include the following:

- Supplementary Movie 1

Supplementary Materials and Methods

Plasma ethanol concentration measurement

Blood samples were collected from the tail snips of AE-control and AE pregnant mice before and 30, 60 min, and one day after intraperitoneal administration of PBS or 25% ethanol solution in PBS on gestational days 16.5 and 17.5. Ethanol concentration was measured by using the Ethanol Assay kit (Abcam, #ab65343). The ethanol concentration was determined using the standard curve, after subtracting the blank value from all standards and samples.

Open field test

The open field test was used to assess the locomotor activity and anxiety-like behavior of the mice. Briefly, C57BL/6J mice were transported to the testing room and allowed to acclimate for 15 min before each test. The testing room was illuminated with overhead lighting at ~450 lux. For the testing session, each mouse was gently placed in a corner of an open field Plexiglas clear chamber (21 cm × 21 cm × 30 cm) and allowed to move freely for 30 min. The data were collected using the open field activity monitoring system (AccuScan Instruments), which uses photocell emitters and receptors forming an x–y grid of invisible infrared beams.

Accelerated rotarod test

Motor learning ability was evaluated with accelerated rotarod test (1). Mice were acclimated with rotarod apparatus for two minutes before testing. Testing consisted of three trials/day, with at least a fifteen-minute break between trials, for two consecutive days (Fig. 1I). Trial was terminated if mice fell off rotarod, or made one complete rotation without walking on rotarod, or reached maximum speed after a five-minute session. Latency to fall from rotarod was recorded automatically by rotarod sensors.

Immunohistochemical quantification of insulin-positive cell areas and islet

Pancreatic tissues were harvested from P78 OMD-controls and OMDs, fixed overnight in 4% PFA, and paraffin-embedded. Paraffin-embedded sections were cut at a thickness of 6 µm, deparaffinized, and were blocked with 5% goat serum for 30 minutes at room temperature. We used a monoclonal rabbit anti-insulin primary antibody (Abcam, ab181547) diluted at 1/200 for immunostaining. Sections were covered and slipped with the prolonged antifade reagent with DAPI to label the nuclei. The entire pancreatic tissue

was scanned at 10x magnification using a fluorescence microscope (BZ-X810, KEYENCE, JAPAN). Labelled area and tissue were quantified by using ImageJ software. The number of islets was counted manually in each tissue section.

PBMC isolation

Blood was drawn from anesthetized mice (Isoflurane, Henry Schein) via cardiac puncture, immediately transferred to collection tubes containing EDTA-2K, and centrifuged at 1500 rpm for five minutes at RT. The top blood plasma layer was discarded. The remnant, containing PBMCs, was mixed with five ml of 10% FBS in DMEM/F12 and Ficoll-Paque PLUS solution (Millipore Sigma), and centrifuged at 1000g for fifteen minutes at RT. The middle layer, containing PBMCs, was collected, washed twice with PBS, and transferred to ice-cold buffer for FACS analysis.

FACS

PBMCs were stained with rat monoclonal anti-mouse CD11b antibody (Cat# 101205, Biolegend), CD19 antibody (Cat# 152407, Biolegend) and CD90.2 antibody (Cat# 105311, Biolegend). Cells were gated based on cell size [forward scatter (FSC) versus side scatter] and singlets (FSC versus trigger pulse width). T-cells, B-cells and monocytes were separated by collecting $CD11b^{+}/CD19^{-}$, $CD19^{+}/CD90.2^{-}$ and $CD19^{-}/CD90.2^{+}$ cell subpopulations respectively.

Library preparation and RNA-sequencing

cDNA libraries were prepared using SMART-Seq v4 Ultra Low Input RNA Sequencing Kit (TAKARA Bio) and Nextera XT DNA Library Prep Kit (Illumina) as per manufacturer instructions, and assessed for their size distribution and concentration using BioAnalyzer High Sensitivity DNA Kit (Agilent Technologies). Libraries were pooled and diluted to 3nM with 10mM Tris-HCl, pH8.5, and denatured as per Illumina protocol. Denatured libraries were loaded onto an S1 flow cell on an Illumina NovaSeq 6000 (Illumina) and ran for 2 x 50 cycles. De-multiplexed sequencing reads were generated using Illumina bcl2fastq (v2.18.0.12) allowing no mismatches in index read.

RNA-seq analysis

Low quality reads were removed via FastQC (sickle with default setting) (2). A HISAT2 index was built for mm10 genome assembly using HISAT2 v2.1.0(3). RNA-sequencing reads of each sample were mapped using HISAT2 or STAR v2.5.4a (4) supplied with

Ensembl annotation file; GRCm38.78.gtf. For count call, HTseq-count v0.10.0 (5) or featurecounts (6) was used.

MDS analysis

MDS analysis of TMM-normalized RNA-sequencing data was performed using IDEAMEX (7).

AS analysis

BAM (binary alignment map) files were used by rMATS (v4.0.2) (8) python script to perform splicing analysis; Number of threads *(--nthread 48*), Gene Transfer File *(--gtf Mus_musculus.GRCm38.78_ERCC92.gtf*), type of read *(-t paired*), the length of each read *(--readLength 50*), the cutoff splicing difference *(--cstat 0.001*), and library type *(--libType fr-unstranded*) were used as parameters.

For differential splicing analysis with Leafcutter (9), BAM files were first converted into JUNC (i.e., junction) files. *Regtools* software was used with these parameters; anchor length *(-a 8)*, minimum length *(-m 50)*, maximum length *(-M 500000)*. List of junc files was generated into *juncfiles.txt.* Then, intron clustering was performed with default parameters using *leafcutter_cluster_regtools.py* python script. Files were then grouped into two datasets for differential splicing analysis (*groups_file.txt)*. Then *leafcutter_ds.R* script was used to generate two files; *leafcutter_ds_cluster_significance.txt* and *leafcutter_ds_effect_sizes.txt*. Lastly spliced junctions were plotted using *ds_plots.R* script.

Quantitative real-time PCR

Quantitative real-time PCR was performed to verify altered expression of spliced transcripts. Test samples were derived from a random subset of samples on which RNA sequencing was performed. The primers and the PCR product length are as follows:

***Ets1*:** Forward - 5’ - AAAAGAACAGCAGCGACTGG - 3’

Reverse - 5’ – TGGGTAGGTAGGGTTGGCTC - 3’

Product length = 280 bp

***Tvp23b*:** Forward - 5’ - AGAGCCACTGGGTGTTTGAG - 3’

Reverse - 5’ – CTGCCCACTTTACACCGGAT - 3’

Product length = 232 bp

GO analysis

GO terms were derived using Enrichr (10). Significant GO terms were analyzed for their semantic similarity, based on their precomputed information content (11), using MDS analysis with REVIGO (12). REVIGO then uses MSDJ, a JAVA library for MDS (v0.2), for MDS analysis (13). Semantic similarity coordinates for each GO term were plotted as scatter plots in Tableau (14) (v2019).

Protein structure prediction and rendering

Three-dimensional structure of proteins were predicted using AlphaFold2 algorithm (Deepmind, Google) (15, 16) along with predicted Local Distance Difference Test (pLDDT) plots and Predicted Aligned Error plots. The predicted protein structures were surface-rendered and annotated with Polyview-3D (17).

Superimposition of protein structures

Three-dimensional protein structures were superimposed using a pairwise structural alignment algorithm, FATCAT (18). RMSD (root mean square distance) for a pair of superimposed structures computes root mean square deviation of aligned Cα atoms of input structures, with one input structure rearranged if flexibility is detected (by introducing twists in the alignment).

RBP binding site density analysis

Chromosome coordinates of regions important for AS of significantly upregulated (+Δpsi) and downregulated (-Δpsi) exons in SE AS, in PAE and OMD, in B-cells and T-cells, were identified via rMAPS2 (19). These coordinates were then used to determine RBP binding site density in these regions via RBPmap (20).

Data presentation resources and tools

Venn diagrams were generated with Biovinci (v2.8.3) (21). Circos plots were generated using R (v3.6) circlize package (v0.4.11) (22). Volcano plots, scatter plots, GO MDS plots, and horizontal bar charts (Fig. 7J-K) were generated with Tableau (v2019) (14).

Tornado plot (Fig. 6A) and LSTM deep-learning model performance plots (Fig. 4B-J, Fig. 5) were generated with Matplotlib (v3.1.0) (23) and Seaborn (v0.10.1) (24) Python libraries. Hierarchical clustering of data into heatmaps with dendrograms was performed with Morpheus (25).

Statistics

Unless otherwise stated, quantitative data is presented as mean with standard error. For data that passed D'Agostino & Pearson normality test, we performed two-tailed Student's *t*-test, one-way or two-way (repeated measures) analysis of variance (ANOVA) followed by post-hoc test (either Holm-Sidak's or Tukey multiple comparison test), as indicated. When statistically significant interaction between independent variables in two-way ANOVA was found, simple main effects were reported. Kruskal-Wallis test was used for nonparametric one-way ANOVA and multiple comparisons. Mann-Whitney U test was used for two-group comparisons of not normally distributed data. GraphPad Prism (v7) was used for statistical analysis. P-values < 0.05 (0.01 for GO analysis) were considered significant.

Supplementary Figs:

Fig. S1

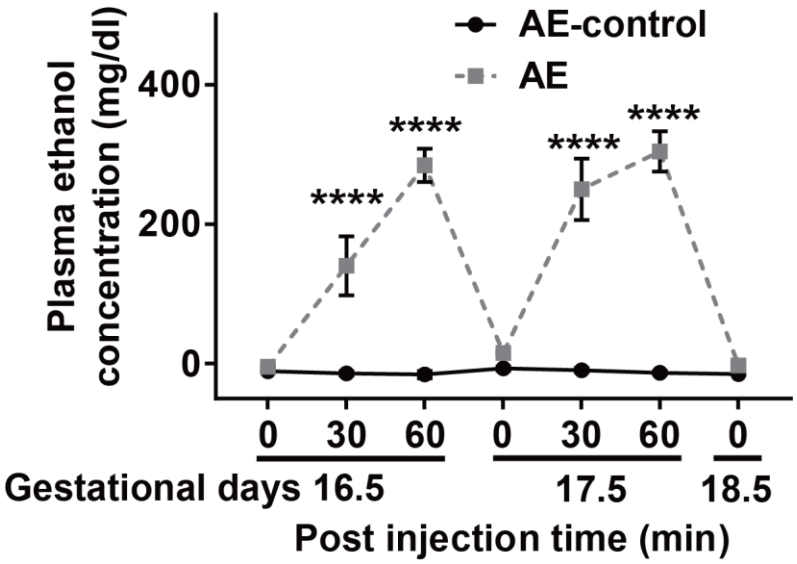


Fig. S1. Dynamics of blood alcohol concentration post administration to pregnant female. AE mice showed higher blood ethanol concentrations at 30 and 60 minutes after receiving ethanol solutions. Treatment x Time interaction: $F(6, 36) = 23.92$, $p < 0.0001$ by two-way repeated measures ANOVA. Sample sizes (# mice): AE-control = 3 mice, AE = 5 mice. Horizontal lines show mean ± SEM. **** $p < 0.0001$ by simple main effect test.

Fig. S2

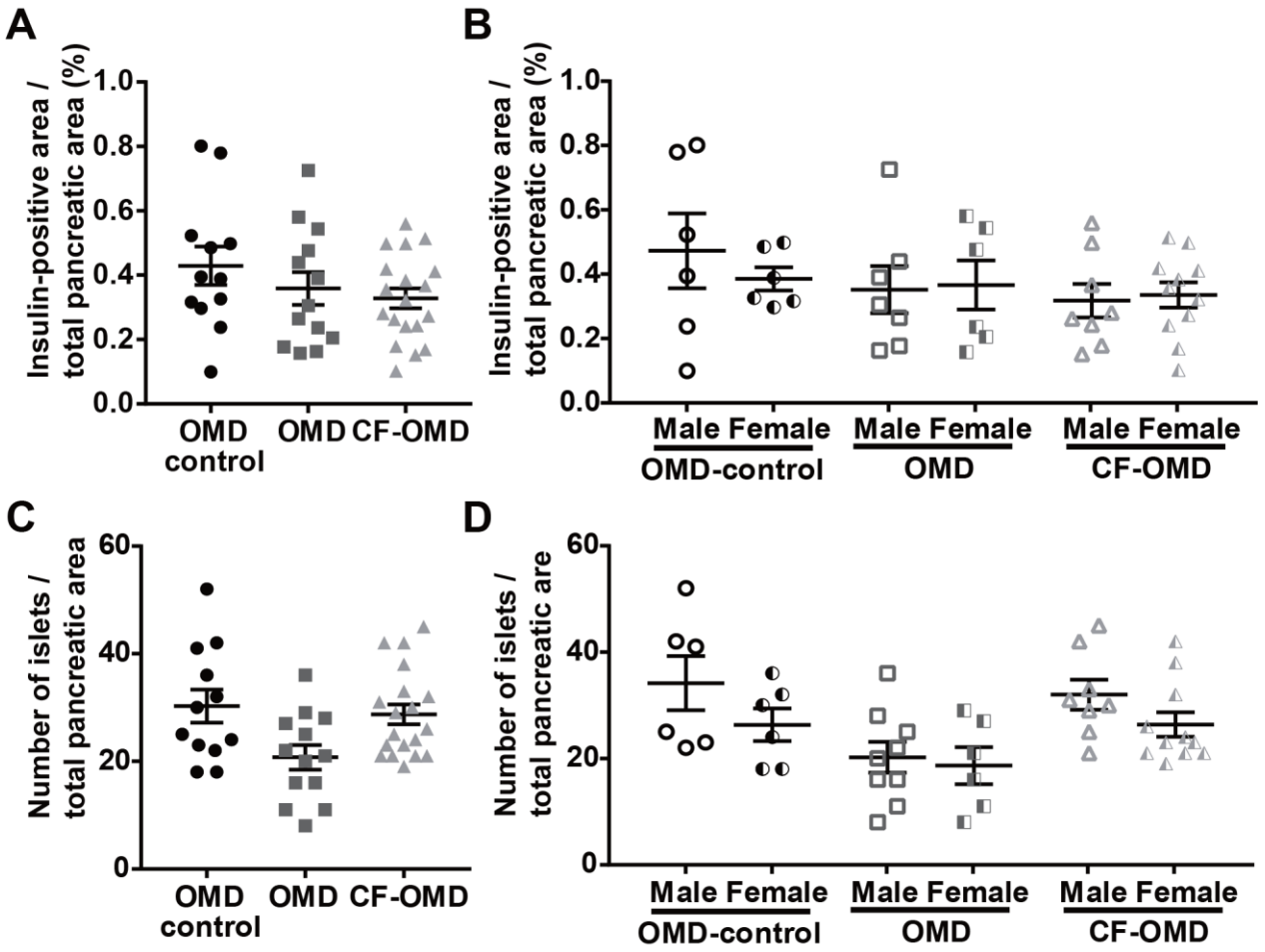


Fig. S2. Normal histology in pancreas from 11-week-old OMD mice. (**A**) Ratio of insulin-positive cell area against entire pancreatic tissue area in OMD-control, OMD, and CF-OMD mice. F (2, 41) = 0.7975, p = 0.4573 by one-way ANOVA. (**B**) No ~~sex dimorphism~~ was observed in this measure [OMD-control male vs OMD-control female, p = 0.2182 by two-tailed Student's t-test; OMD male vs OMD female, p = 0.7376 by two-tailed Student's t-test; CF-OMD male vs CF-OMD female, p = 0.1019 by Mann-Whitney U test. (**C**) The number of Islet cells is similar between these three groups (Kruskal-Wallis statistic = 6.664], and (**D**) between gender-divided sub-groups (MD-control male vs OMD-control female, p = 0.4886; OMD male vs OMD female, p = 0.8943; CF-OMD male vs CF-OMD female, p = 0.7846 by two-tailed Student's t-test). Sample sizes: (A-D) OMD-control = 12 mice (6 males), OMD = 13 (7), CF-OMD = 19 (8). Horizontal lines represent mean ± SEM.

Fig. S3

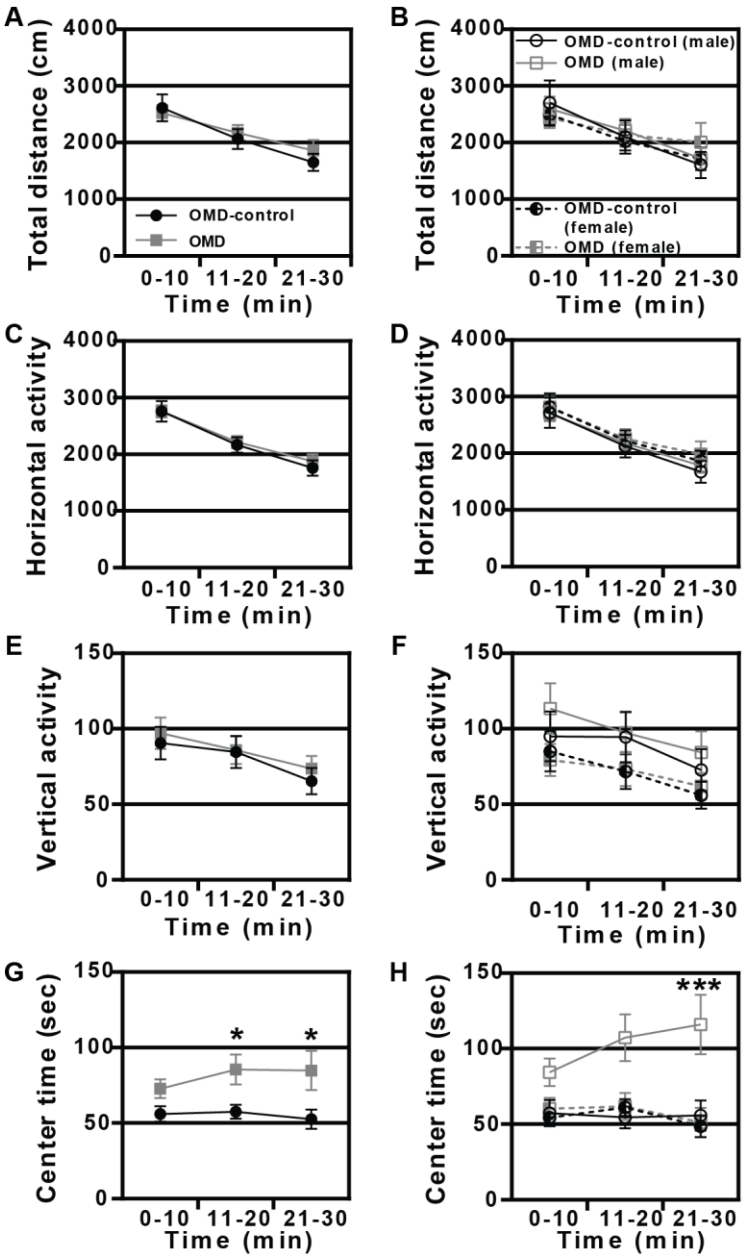


Fig. S3. OMD exhibit anxiety-like behavior while maintaining normal locomotion in the open field test. (**A**-**F**) There was no difference in total distance (A, B), horizontal activity (C, D), or vertical activity (E, F) between the OMD and the OMD-control group (A, C, E) [A; Treatment: $F(1, 55) = 0.09975$, $p = 0.7533$, Days: $F(2, 110) = 43.77$, $p < 0.0001$, Interaction: $F(2, 110) = 1.527$, $p = 0.2218$. C; Treatment: $F(1, 55) = 0.09666$, $p = 0.7571$, Days: $F(2, 110) = 113.4$, $p < 0.0001$, Interaction: $F(2, 110) = 0.4602$, $p = 0.6324$. E; Treatment: $F(1, 55) = 0.1532$, $p = 0.6979$, Days: $F(2, 110) = 26.48$, $p < 0.0001$, Interaction: $F(2, 110) = 0.5591$, $p = 0.5733$.] or between gender separated sub-groups (B, D, F) [B;

Treatment: F (3, 53) = 0.4345, p = 0.9878, Days: F (2, 106) = 42.41, p < 0.0001, Interaction: F (6, 106) = 1.387, p = 0.2265. D; Treatment: F (3, 53) = 0.1879, p = 0.9042, Days: F (2, 106) = 108.5, p < 0.0001, Interaction: F (6, 106) = 0.2801, p = 0.9452. F; Treatment: F (3, 53) = 0.9083, p = 0.4433, Days: F (2, 106) = 26.30, p < 0.0001, Interaction: F (6, 106) = 0.8003, p = 0.5718.]. No significant difference in comparisons at each time point by two-tailed Student's t-test. (G, H) OMD group (G), especially OMD males (H), spent significantly more time in the center area compared to control group [G; Treatment: F (1, 55) = 7.109, p = 0.01000, Days: F (2, 110) = 0.2864, p = 0.2864, Interaction: F (2, 110) = 1.625, p = 0.2015. H; Treatment x Days interaction: F (6, 106) = 2.530, p = 0.02500.]. Statistical tests: (A, C, E, G) two-way repeated measures ANOVA followed by Holm-Sidak's multiple comparisons test. (B, D, F) two-way repeated measures ANOVA followed by Tukey's multiple comparisons test. (H) Two-way repeated measures ANOVA, followed by simple main effect test. OMD-control male = 18 mice, OMD male = 13, OMD-control female = 14, OMD female = 12. Graphs show mean ± SEM. * p < 0.05. *** p < 0.001.

Fig. S4

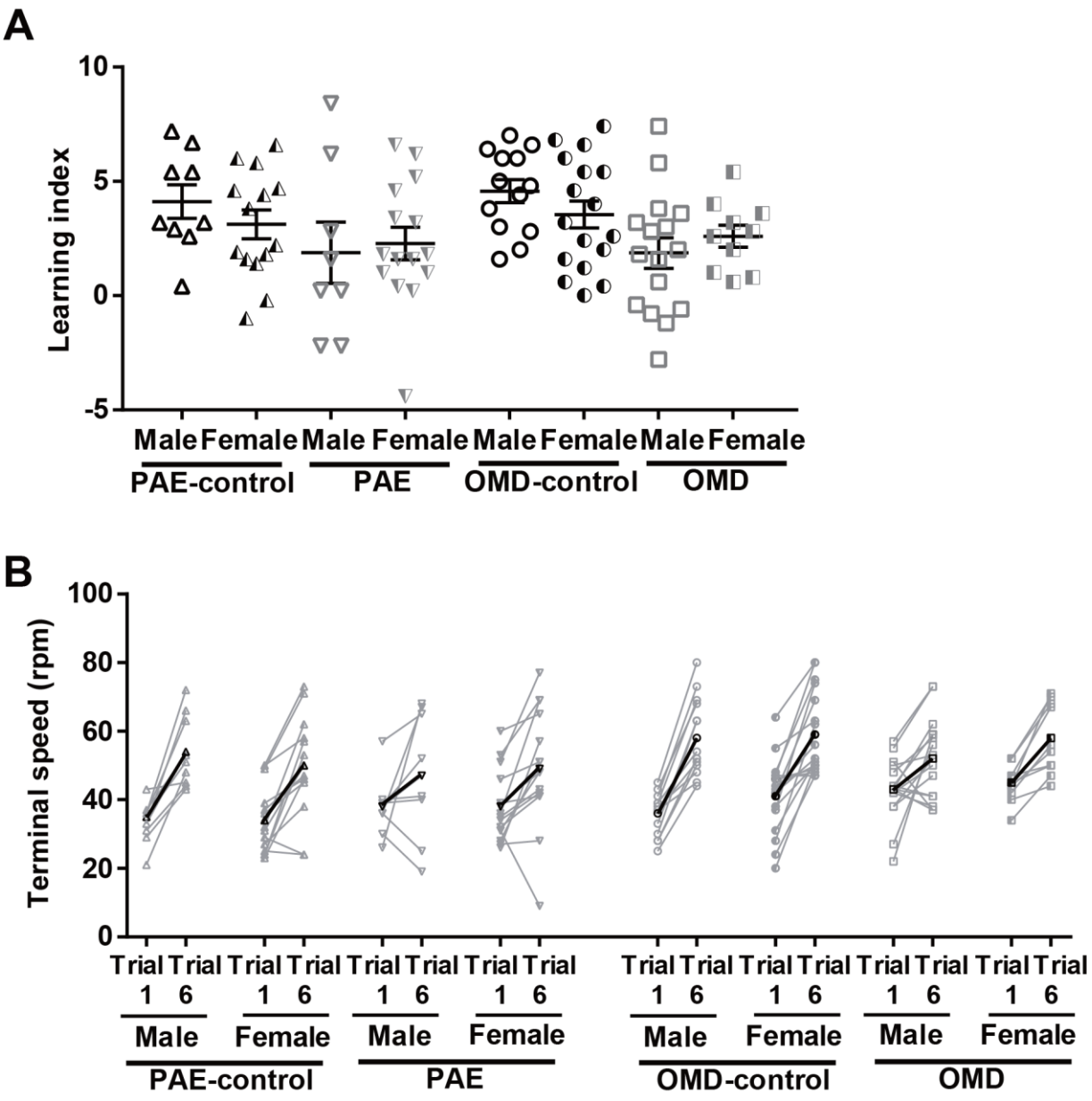


Fig. S4. Gender-Based Analysis of Motor Learning. (**A**) Learning index is not different between male and female in each experimental group [PAE-control male vs PAE-control female, p = 0.3211; PAE male vs PAE female, p = 0.7718; OMD-control male vs OMD-control female, p = 0.9032; OMD male vs OMD female, p = 0.4384]. Graph presents mean ± SEM. (**B**) Changes in terminal speed from trial 1 to 6 are similar between male and female in each experimental group [PAE-control male vs PAE-control female, p = 0.3206; PAE male vs PAE female, p = 0.7718; OMD-control male vs OMD-control female, p = 0.2118; OMD male vs OMD female, p = 0.4384]. Dense lines show mean value. Two-tailed

Student’s t-test. (A, B) PAE-control male = 9 mice, PAE male = 8, PAE-control female = 14, PAE female = 15, OMD-control male = 13, OMD male = 16, OMD-control female = 17, OMD female = 10.

Fig. S5

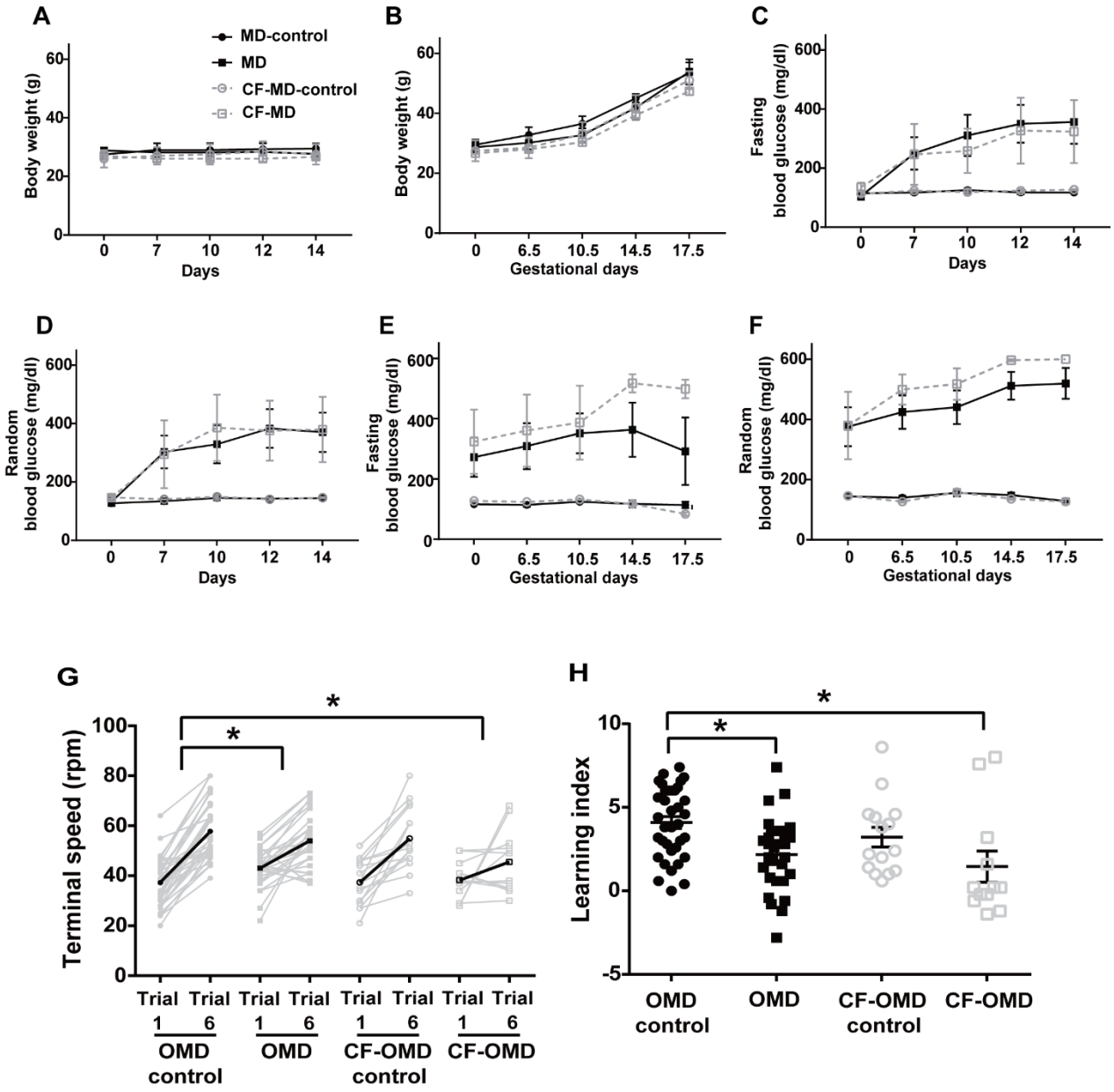


Fig. S5. Cross-fostering neither changes mothers' metabolic profiles post STZ treatment or motor skill learning in OMD mice. (**A**, **B**) There was no difference in body weight between mother with diabetes (MD) and cross-fostering mother with diabetes (CF-MD), and between MD-control (mother injected with CB instead of STZ) and cross fostering control mother (CF-MD-control) at indicated days post injection of STZ (or CB, for control mice) (A) [MD vs CF-MD; Cross-fostering: $F(1, 5) = 6.129$, $p = 0.0561$, Days: $F(4, 20) = 0.9635$, $p = 0.4490$, Interaction: $F(4, 20) = 0.6158$, $p = 0.6563$, MD-control vs CF-MD-control; Cross-fostering: $F(1, 3) = 0.02424$, $p = 0.8862$, Days: $F(4, 12) = 5.077$,

p = 0.0125, Interaction: F (4, 12) = 0.8191, p = 0.5373], and post conception (B) [MD vs CF-MD; Cross-fostering: F (1, 5) = 3.271, p = 0.1303, Days: F (4, 20) = 94.19, p < 0.0001, Interaction: F (4, 20) = 2.588, p = 0.0681, MD-control vs CF-MD-control; Cross-fostering: F (1, 3) = 0.5219, p = 0.5223, Days: F (4, 12) = 72.97, p < 0.0001, Interaction: F (4, 12) = 0.1912, p = 0.9384]. (**C**, **D**) There was no difference in blood glucose levels between MD and CF-MD mice, and between MD-control and CF-MD-control mice at indicated days post injection of STZ (or CB, for control mice). (C) [MD vs CF-MD; Cross-fostering: F (1, 5) = 0.4828, p = 0.5181, Days: F (4, 20) = 7.327, p = 0.0008, Interaction: F (4, 20) = 0.07534, p = 0.9889, MD-control vs CF-MD-control; Cross-fostering: F (1, 3) = 2.003, p = 0.2519, Days: F (4, 12) = 0.6055, p = 0.6662, Interaction: F (4, 12) = 1.32, p = 0.3178] and sampling at random intervals (D) [MD vs CF-MD; Cross-fostering: F (1, 5) = 0.8174, p = 0.4074, Days: F (4, 20) = 0.4421, p = 0.7768, Interaction: F (4, 20) = 0.4421, p = 0.7768, MD-control vs CF-MD-control; Cross-fostering: F (1, 3) = 6.568, p = 0.0830, Days: F (4, 12) = 8.113, p = 0.0005, Interaction: F (4, 12) = 0.512, p = 0.7284]. (**E**, **F**) There was no difference in blood glucose level between MD and CF-MD mice, and between MD-control and CF-MD-control mice at the indicated days post conception. (E) [MD vs CF-MD; Cross-fostering: F (1, 5) = 0.8109, p = 0.4091, Days: F (4, 20) = 2.786, p = 0.0546, Interaction: F (4, 20) = 1.347, p = 0.2874, MD-control vs CF-MD-control; Cross-fostering: F (1, 3) = 0.002431, p = 0.9638, Days: F (4, 12) = 10.39, p = 0.0007, Interaction: F (4, 12) = 6.27, p = 0.0058] and sampling time (F) [MD vs CF-MD; Cross-fostering: F (1, 5) = 2.872, p = 0.1509, Days: F (4, 20) = 7.842, p = 0.0006, Interaction: F (4, 20) = 0.1697, p = 0.9513, MD-control vs CF-MD-control; Cross-fostering: F (1, 3) = 0.06531, p = 0.8148, Days: F (4, 12) = 7.912, p = 0.0023, Interaction: F (4, 12) = 1.788, p = 0.1960]. (**G**, **H**) Similar to OMD mice, cross-fostered OMD (CF-OMD) mice lower performance in the accelerated rotarod motor test (G) [F (3, 79) = 4.569, p = 0.0050] and had a significantly lower learning index than OMD-control mice (H) [F (3, 79) = 4.405, p = 0.0064 by one-way ANOVA. *p < 0.05 by Holm-Sidak's multiple comparisons test]. Statistical tests: (A-F) Two-way repeated measures ANOVA. (G, H) One-way ANOVA, followed by Tukey’s (G) or Holm-Sidak's multiple comparisons test (H). Fasting blood glucose was determined by collecting blood post 6 hours fasting. (A-F) MD = 4 mice, MD-control = 3, CF-MD-control = 2, CF-MD = 3. (G, H) OMD-control = 30 mice, OMD = 26, CF-OMD-control = 15, CF-OMD = 12. Graphs present mean ± SEM. *p < 0.05.

Fig. S6

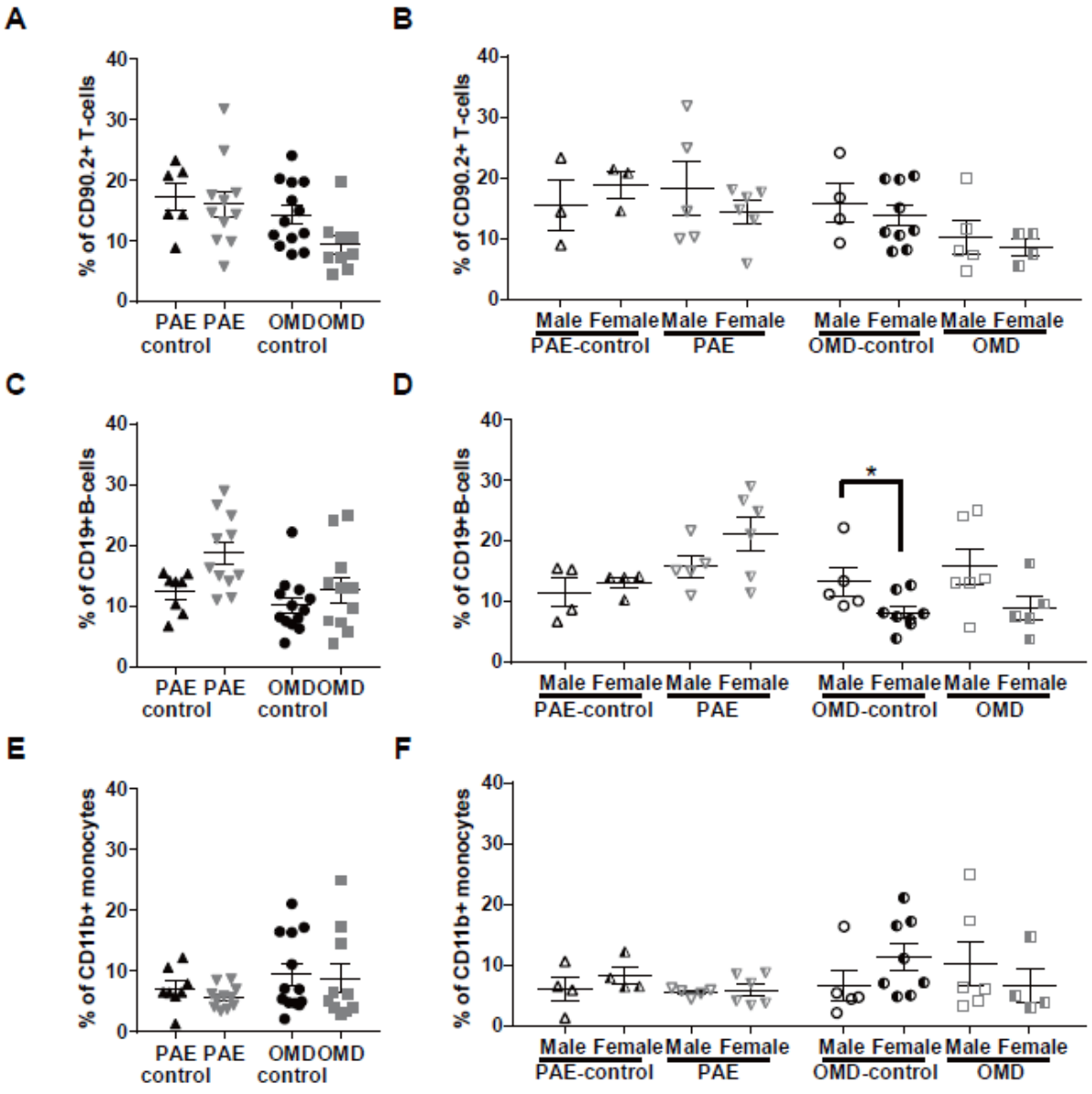


Fig. S6. The proportion of T-cells, B-cells and monocytes in peripheral blood mononuclear cells. (**A**, **C**, **E**) The proportions of T-cells (A), B-cells (C), and monocytes (E) are similar between the respective groups of PAE, OMD, and their control groups; (A) [F (3, 35) = 2.927, p = 0.0472], (C) [F (3, 39) = 5.072, p = 0.0046], (E) [F (3, 38) = 1.132. p = 0.3485]. (**B**, **D**, **F**) The proportions of T-cells (B) and monocytes (F) are not different between male and female in each experimental group. OMD-control female shows a decrease in the proportion of B-cells compared to male. No differences were observed between males and females in all other comparisons. (B) [PAE-control male vs PAE-control female, p = 0.5168; PAE male vs PAE female, p = 0.3987; OMD-control male vs OMD-control female, p = 0.5395; OMD male vs OMD female, p = 0.6211], (D) [PAE-control male vs PAE-control female, p > 0.9999; PAE male vs PAE female, p = 0.1646; OMD-control male vs OMD-control female, p = 0.0452; OMD male vs OMD female, p = 0.1078],

(F) [PAE-control male vs PAE-control female, p = 0.4000; PAE male vs PAE female, p = 0.7345; OMD-control male vs OMD-control female, p = 0.2067; OMD male vs OMD female, p = 0.4721]. Statistical tests: (A, C, E) One-way ANOVA, followed by Tukey's test. (B) two-tailed Student's t-test. (D, F) PAE-control male vs PAE-control female; Mann-Whitney U test, followed by Dunn's multiple comparisons test, PAE male vs PAE female, OMD-control male vs OMD-control female, OMD male vs OMD female; two-tailed Student's t-test. (A) PAE-control = 6 mice, PAE = 11, OMD-control = 13, OMD = 9. (B) Males: PAE-control =3 mice, PAE = 5, OMD-control = 4, OMD = 5, Females: PAE-control = 3 mice, PAE = 6, OMD-control = 9, OMD = 4. (C) PAE-control = 8 mice, PAE = 11, OMD-control = 13, OMD = 9. (D) Males: PAE-control = 4 mice, PAE = 5, OMD-control = 5, OMD = 6. Females: PAE-control = 4 mice, PAE = 6, OMD-control = 8, OMD = 5 mice. (E) PAE-control = 8 mice, PAE = 11, OMD-control = 13, OMD = 10. (F) Males: PAE-control = 4 mice, PAE = 5, OMD-control = 5, OMD = 6 mice. Females: PAE-control = 4 mice, PAE = 6, OMD-control = 5, OMD = 4. Graphs represent mean ± SEM. *$p < 0.05$.

Fig. S7

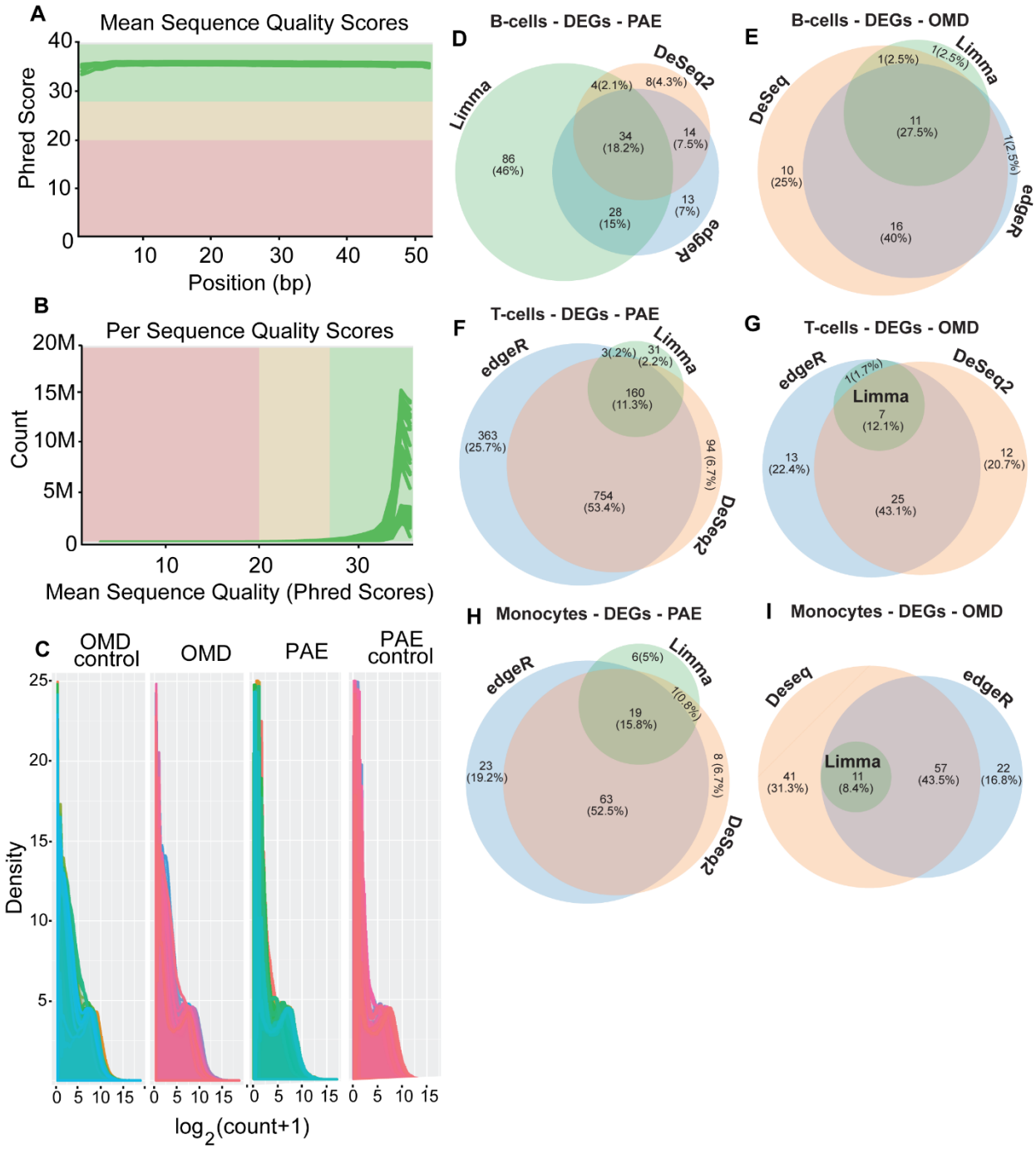


Fig. S7. Quality control evaluations and differential gene expression analysis of RNA sequencing raw data from peripheral immune cells of PAE and OMD mice. (**A**) High Phred (mean sequence quality) scores were observed across all reads. Scores from all reads were averaged per sample, and plotted for all samples. (**B**) The number of read counts in each Phred score are shown. Each line represents one sample. (**C**) Density plots show similar pattern in gene numbers with the indicated number of read counts across samples. (**D**-**G**) Venn Diagrams show the number and proportion of unique and common

significant differentially expressed genes (DEGs) in B-cells (D, E), T-cells (F, G), and monocytes (H, I) in PAE (D, F, H) and OMD (E, G, I) mice as assessed by Limma, DeSeq2 and EdgeR methods of differential gene expression analysis. At least a two-fold change in expression, either up or down, was considered significant. DEGs identified by EdgeR were selected for downstream analyses.

Fig. S8

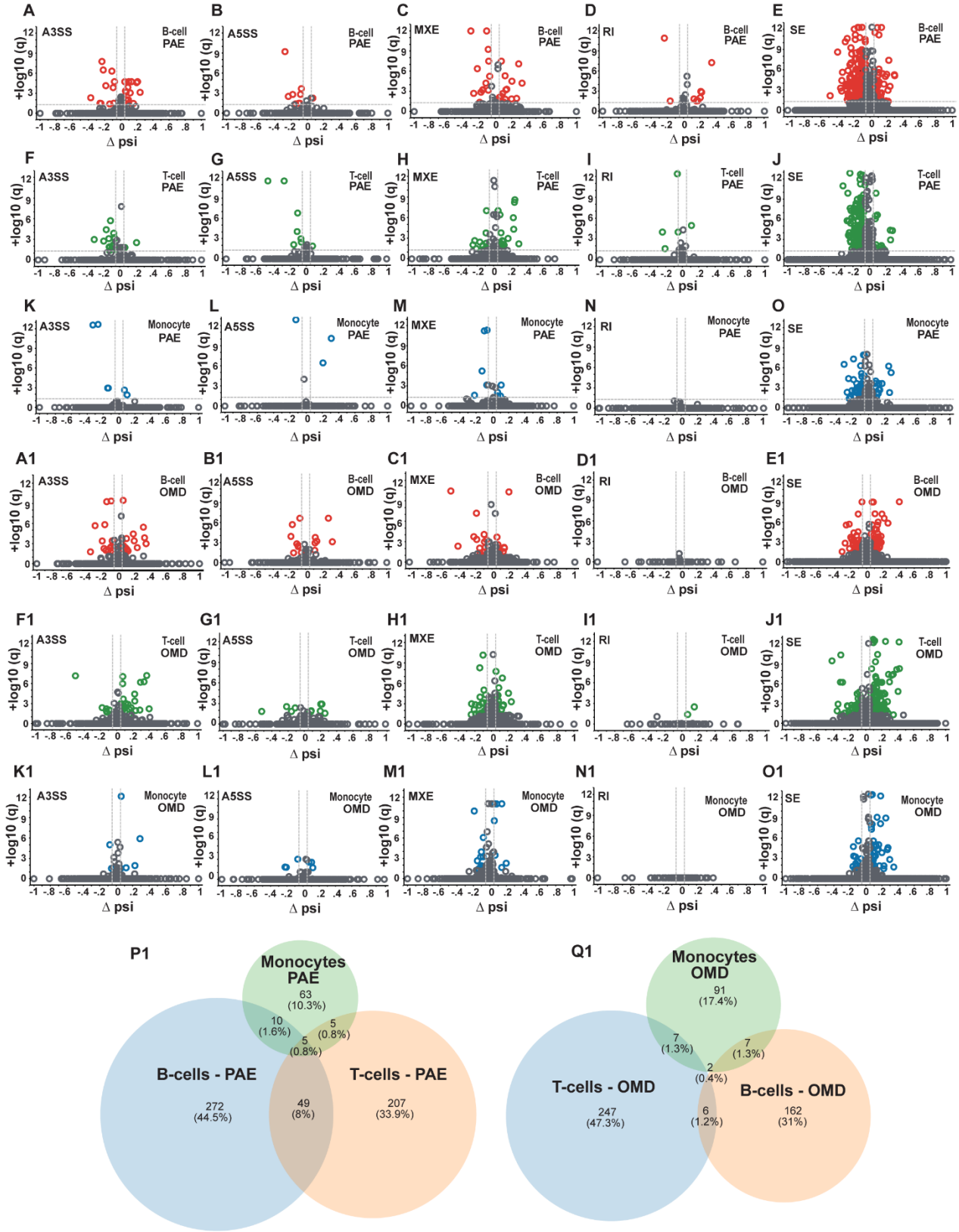


Fig. S8. Distribution of significance and Δpsi values of alternative splicing events in peripheral immune cells of PAE and OMD mice. (**A**-**O1**) Volcano plots show the significance (False Discovery Rate, q-value) and Δpsi values of all AS events identified in

B-cells (A-E, A1-E1), T-cells (F-J, F1-J1), and monocytes (K-O, K1-O1) in PAE (A-O) and OMD (A1-O1) mice. All 5 types of AS events; A3SS (A, F, K, A1, F1, K1), A5SS (B, G, L, B1, G1, L1), MXE (C, H, M, C1, H1, M1), RI (D, I, N, D1, I1, N1) and SE (E, J, O, E1, J1, O1); are represented. (**P1**, **Q1**) Venn diagrams show the number and proportion of unique and common AS events among B-cells, T-cells, and monocytes in PAE (P1) and OMD (Q1) conditions.

Fig. S9

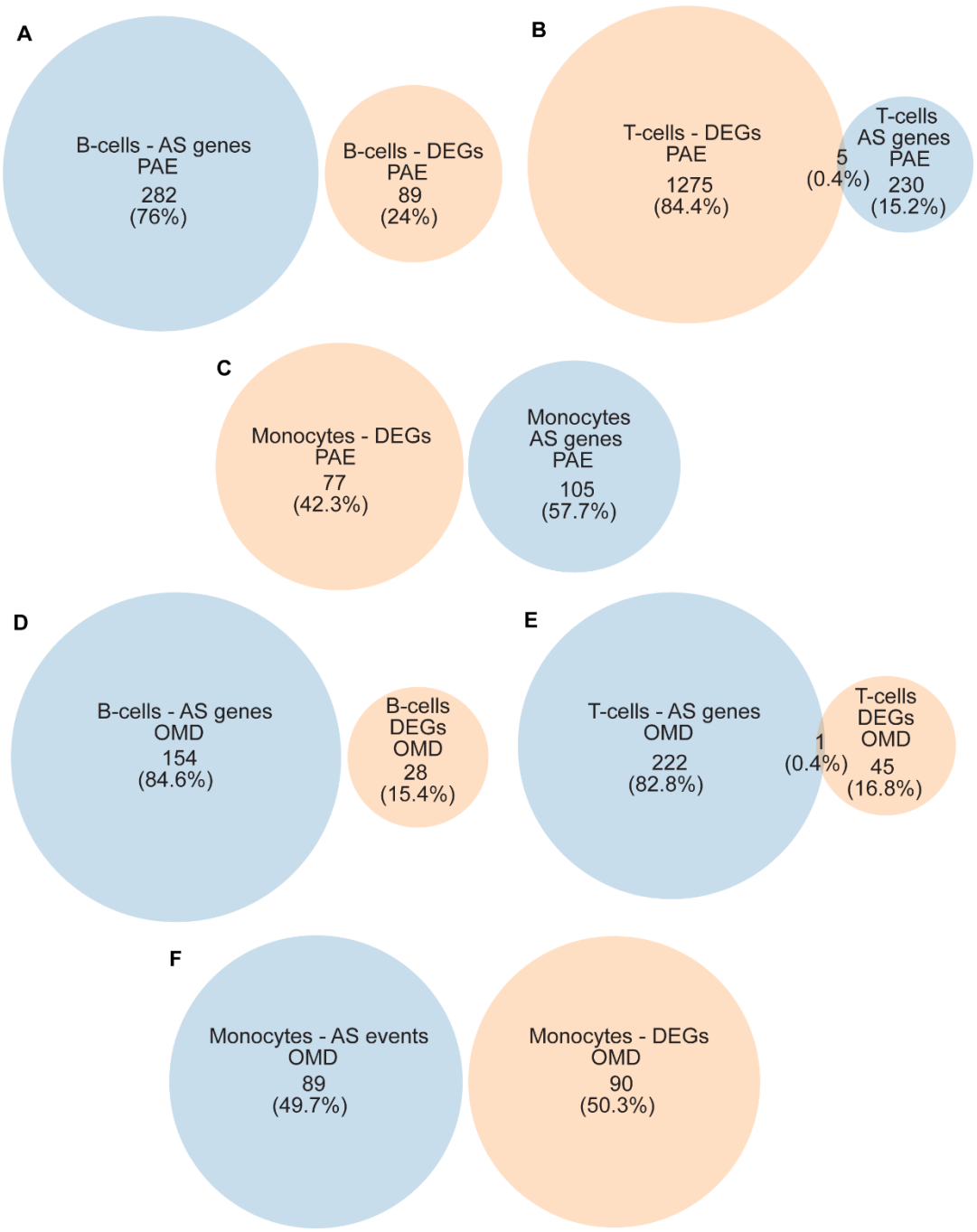


Fig. S9. Minimal to no overlap between DEGs and alternatively spliced genes in peripheral immune cells of PAE and OMD mice. (**A**-**F**) Venn diagrams show the number and proportion of unique and common genes that were significantly alternatively spliced and significantly differentially expressed in B-cells (A, D), T-cells (B, E), and monocytes (C, F) in both PAE (A-C) and OMD (D-F) mice. At least a two-fold change in gene expression, either up or down, as identified by EdgeR analysis, was considered significant. Genes with AS events with a delta psi greater than 5% and a false discovery rate smaller than 5% were considered significantly alternatively spliced.

Fig. S10

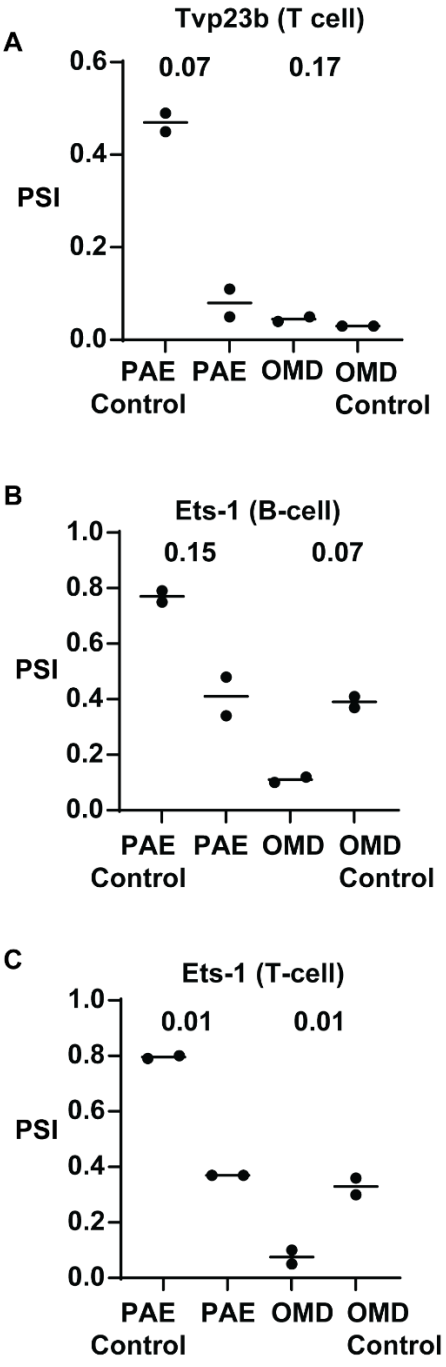


Fig. S10. Quantitative real time PCR validation of differential alternative splicing of key AS events. (A-C) Psi values were calculated from results obtained via quantitative real time PCR for *Tvp23b* gene (A) and *Ets1* (B) gene in B-cells, and *Ets1* gene (C) in T-cells for PAE, PAE-control, OMD and OMD-control groups. P-values from two-tailed Student's t-test are shown for each comparison. N = 2 samples per group.

Fig. S11

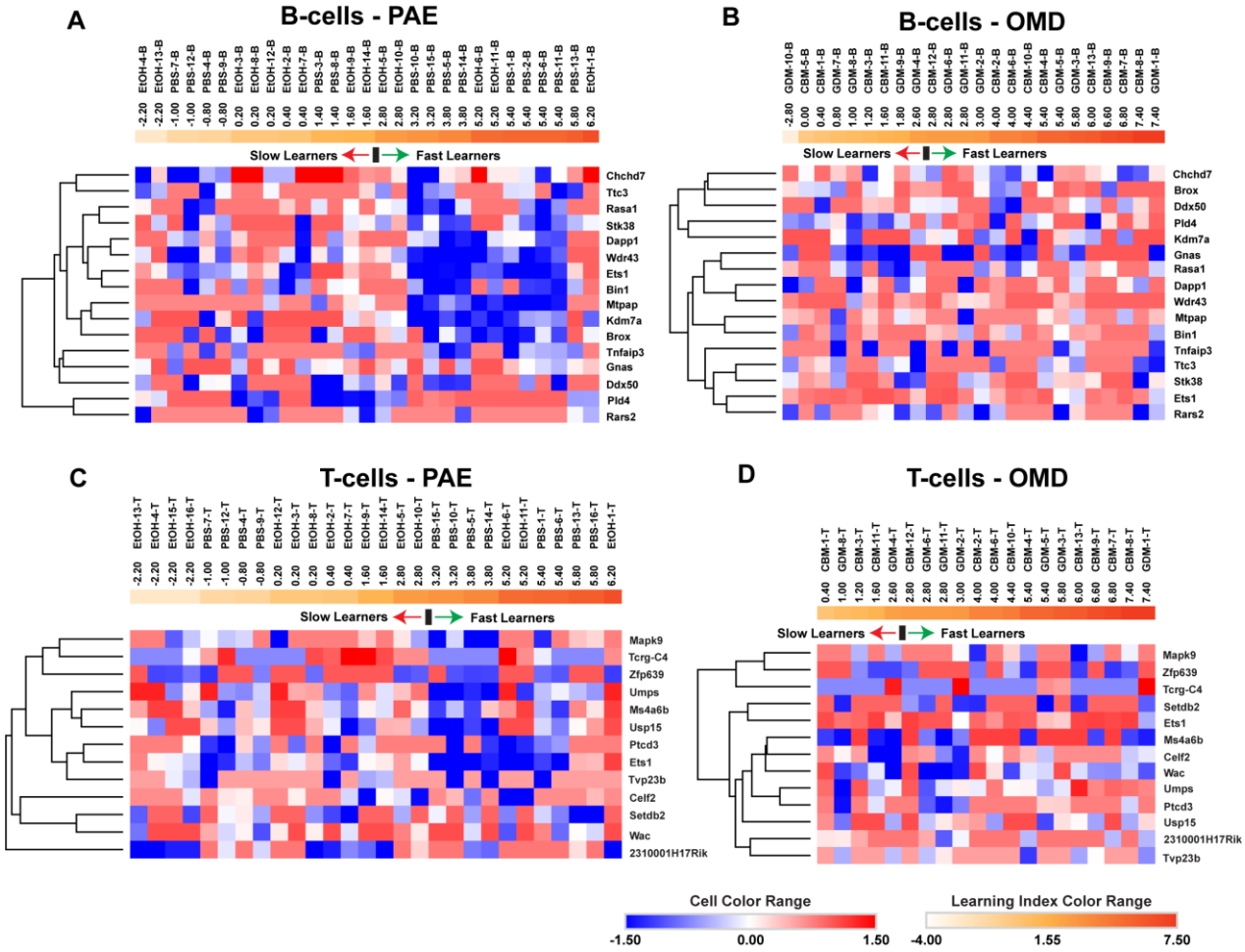


Fig. S11. Clustering of normalized psi values of biomarker AS events according to learning index scores of PAE and OMD mice. (**A**-**D**) Heatmaps show the clustering of z-score normalized psi values of B-cell (A, B) and T-cell (C, D) biomarker AS events from mice in the PAE (A, C) and OMD (B, D) conditions arranged in ascending order of their learning index. The median learning index score of 2.8 is used for binary classification of mice into fast and slow learners. Hierarchical clustering of biomarkers is represented as dendrograms to the left of each heatmap. The color range of heatmap cells are indicated at the bottom. The distribution of slow and fast learners in different sub-groups: B-cells – PAE group: 15 slow learners, 14 fast learners; B-cells – OMD group: 9 slow learners, 15 fast learners; T-cells – PAE group: 17 slow learners, 11 fast learners; and T-cells – OMD group: 5 slow learners, 15 fast learners.

Fig. S12

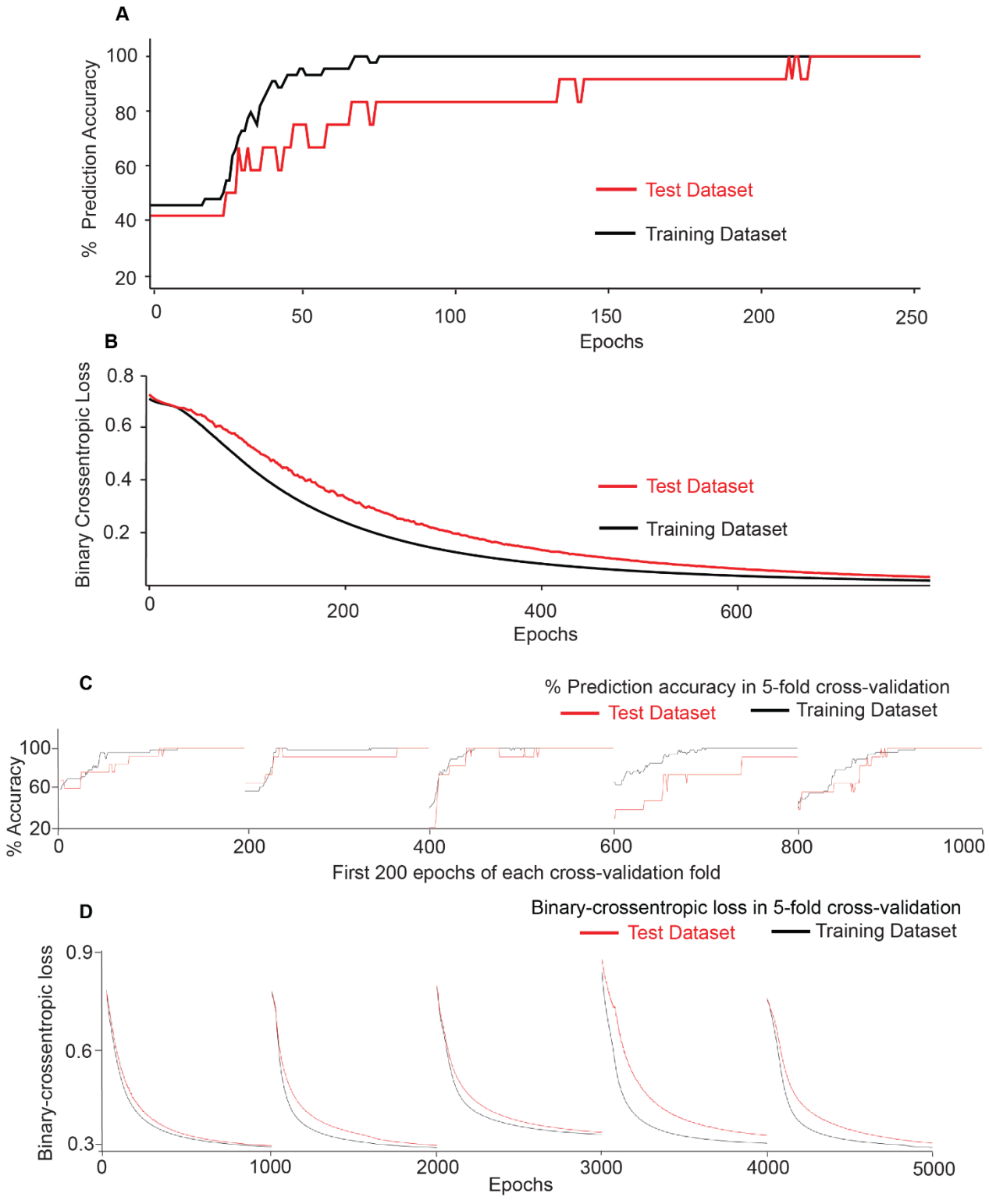


Fig. S12. Aptly trained deep learning model predicts motor learning ability of PAE and OMD mice from psi values of common AS event biomarkers. (**A**) Accuracy of LSTM model in predicting learner type (fast or slow), with each consecutive epoch, in training (black) and test (red) datasets, is plotted as %prediction accuracy over epochs.

Absence of overperformance in test dataset, relative to training dataset, over all epochs, demonstrates that LSTM model is not overfit. (**B**) Binary cross-entropic loss values are plotted for training (black) and test (red) datasets for each epoch. Absence of better learning in test dataset, relative to training dataset, over all epochs, demonstrates that LSTM model is not overfit. Negative slope of loss function curves in training and testing datasets, with their respective loss function values trending towards zero with each successive epoch, demonstrates that LSTM model is not underfit. (**C-D**) Five-fold cross validation of LSTM model. Input dataset was randomly split into five equal parts. During each cross validation; 0-1000, 1000-2000, 2000-3000, 3000-4000, 4000-5000 epochs; any four of five splits were randomly selected to be training dataset and the remaining fifth split was used as test dataset, to evaluate model performance [% prediction accuracy (C)] and model learnability [binary cross-entropic loss (D)]. Black line = Training dataset. Red line = Test dataset.

Fig. S13

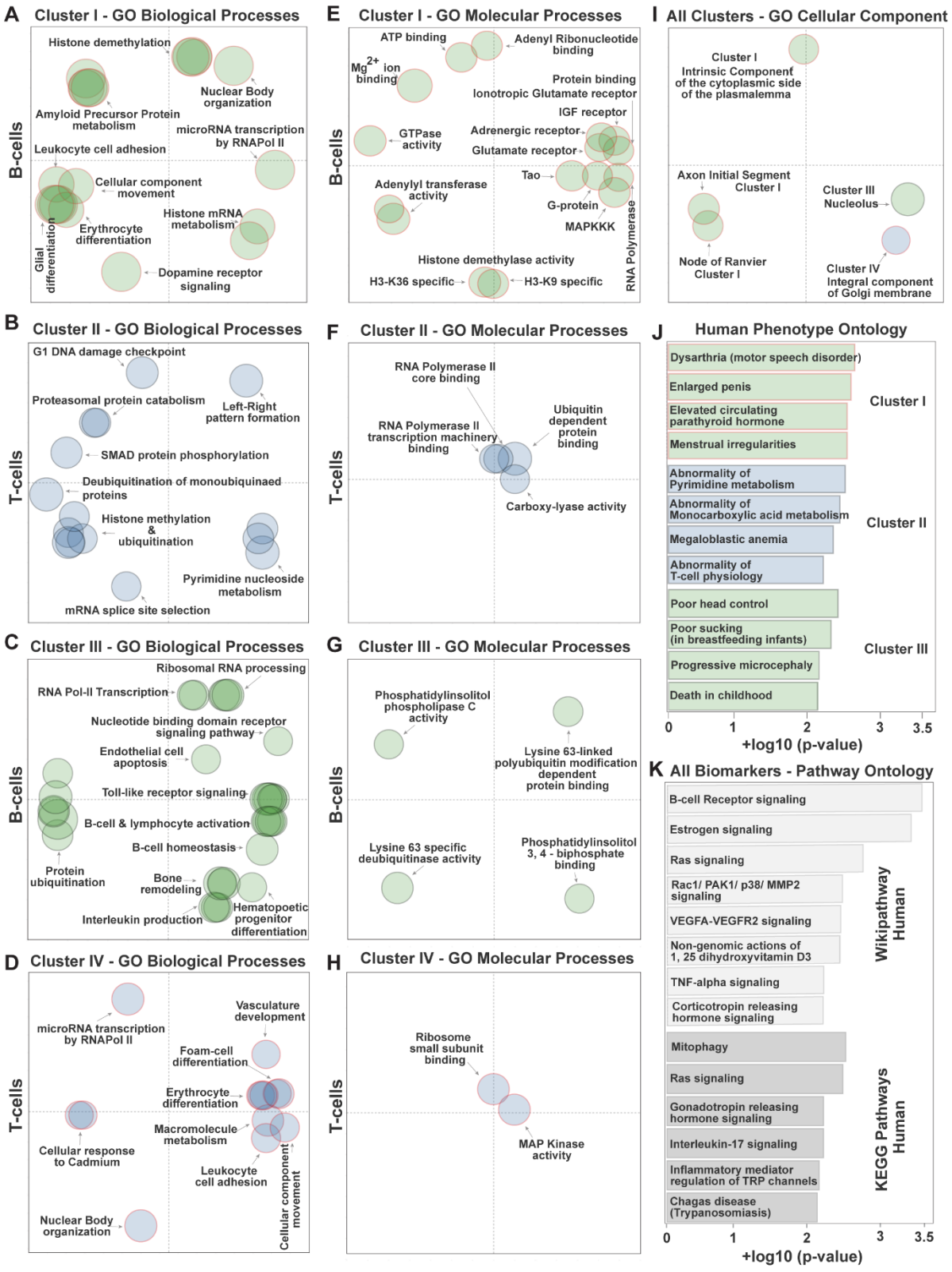


Fig. S13. GO analysis of biomarkers in Shapley-value clusters reveals their relevance for optimal neural function and motor behavior. MDS plots of significant GO

terms for genes within clusters I-IV. (**A**-**D**) MDS plots of significant GO terms for biological processes in cluster I B-cell (A), cluster II T-cell (B), cluster III B-cell (C), and cluster IV T-cell (D) genes. (**E**-**H**) MDS plots of significant GO terms for molecular processes in cluster I (E), cluster II (F), cluster III (G), and cluster IV (H) genes. (**I**) MDS plots of significant GO terms for cellular components in all clusters. (**J**) Human phenotype ontology GO terms for clusters I-III. (**K**) Significant pathway ontology GO terms from Human Wiki and KEGG Pathway databases. GO terms with p-value < 0.01 were considered significant.

Fig. S14

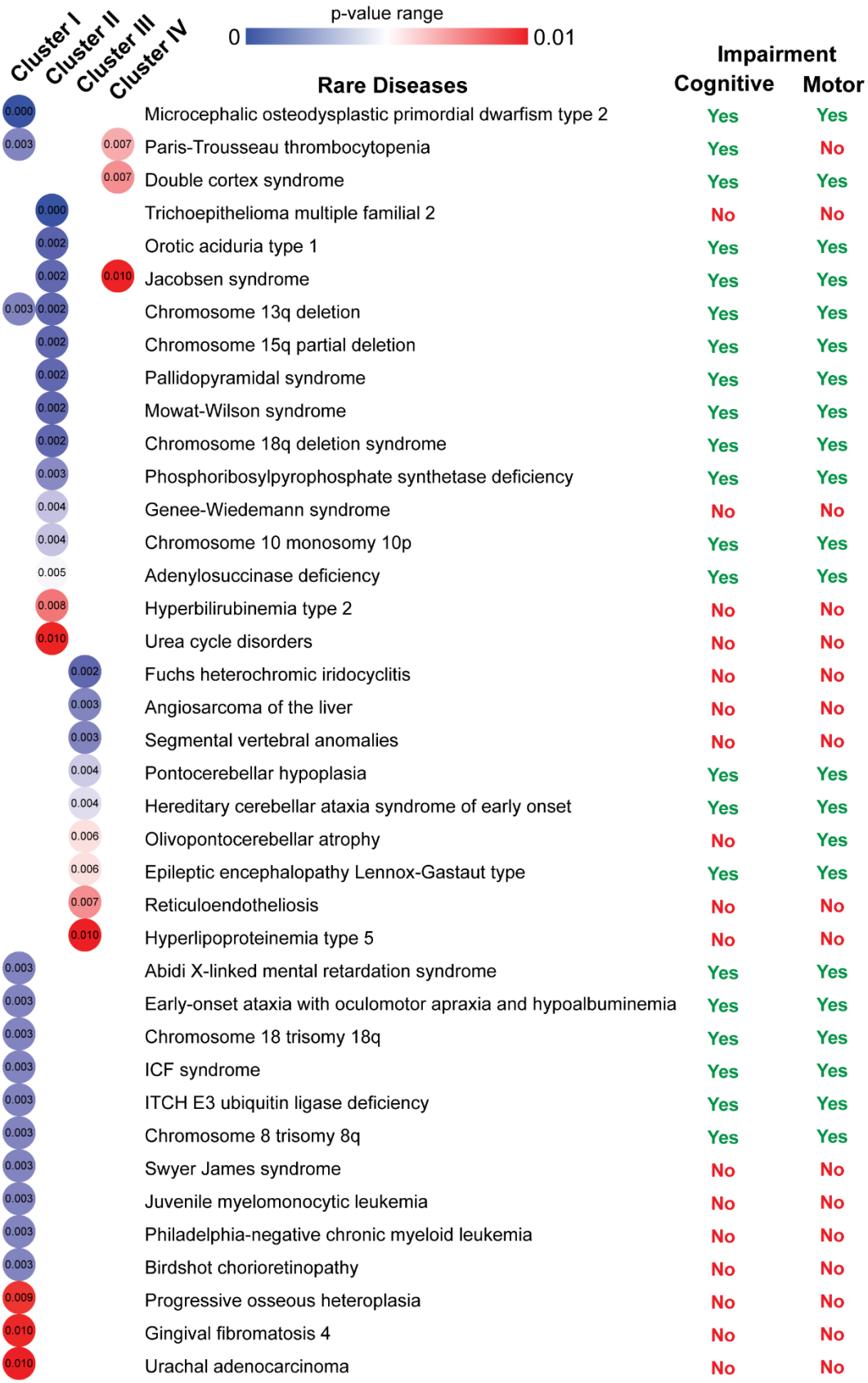
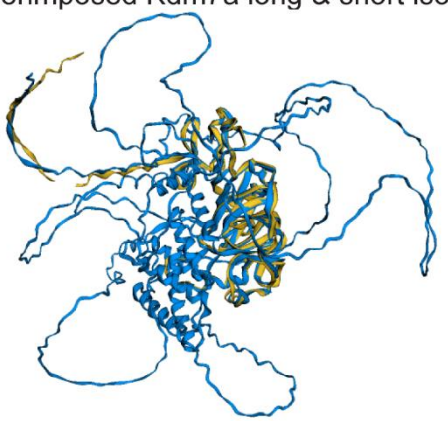


Fig. S14. Rare disease GO analysis of Shapley value-correlated gene clusters reveals relevance of the common AS biomarkers for cognitive and motor

impairment in humans. Significant rare disease GO terms significantly associated with Shapley value-correlated gene clusters I-IV are listed as a table. Each GO term was annotated for the presence (yes) or absence (no) of cognitive and motor impairment on the right. The significance of each GO term is depicted as a colored bubble on a scale of 0 (blue) to 0.01 (red) on the left with the p-value annotated inside of it. Within each cluster, the GO terms are arranged in descending order of their significance. GO terms with a p-value smaller than 0.01 were considered significant. The color range of p-value bubbles is indicated in the top.

Fig. S15

A Superimposed Kdm7a long & short isoform with 1 twist

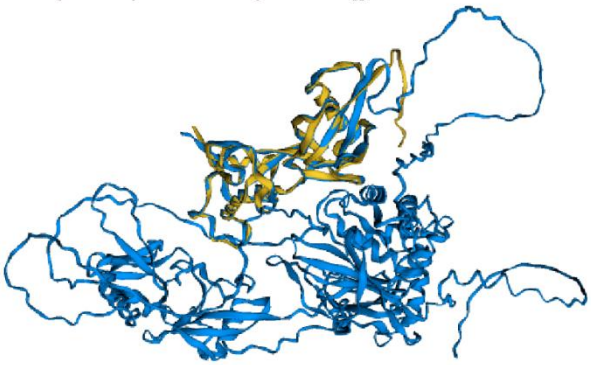


Kdm7a long isoform in blue

Kdm7a short isoform in yellow

RMSD of equivalent positions = 1.42Å

B Superimposed Usp15 long & short isoform with no twist

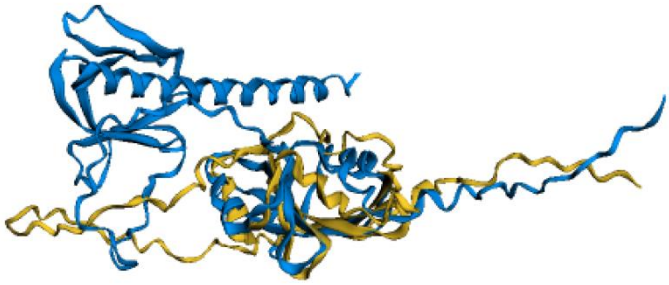


Usp15 long isoform in blue

Usp15 short isoform in yellow

RMSD of equivalent positions = 2.10Å

C Superimposed Dapp1 long & short isoform with 2 twists

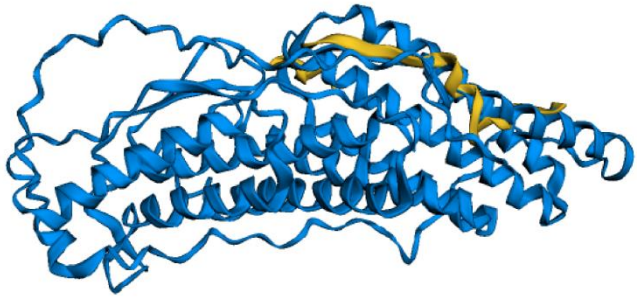


Dapp1 long isoform in blue

Dapp1 short isoform in yellow

RMSD of equivalent positions = 3.77Å

D Superimposed Brox long & short isoform with no twist



Brox long isoform in blue

Brox short isoform in yellow

RMSD of equivalent positions = 3.34Å

Fig. S15. Superimposition of predicted structures of selected biomarkers of interest show gross mismatch between their long and short isoforms. (**A-D**) Superimposition of AlphaFold2 predicted structures of the long (blue, A-D) and short (yellow, A-D) alternatively spliced isoforms of Kdm7a (A), Usp15 (B), Dapp1 (C) and Brox (D) with 1, 0, 2, and 0 twists respectively. In each case, there is a gross mismatch between the short and long protein isoforms. The RMSD of aligned Cα atoms between the long and short isoforms of Kdm7a, Usp15, Dapp1, and Brox were 1.42Å (A), 2.10Å (B), 3.77Å (C), and 3.34Å (D), respectively.

Fig. S16

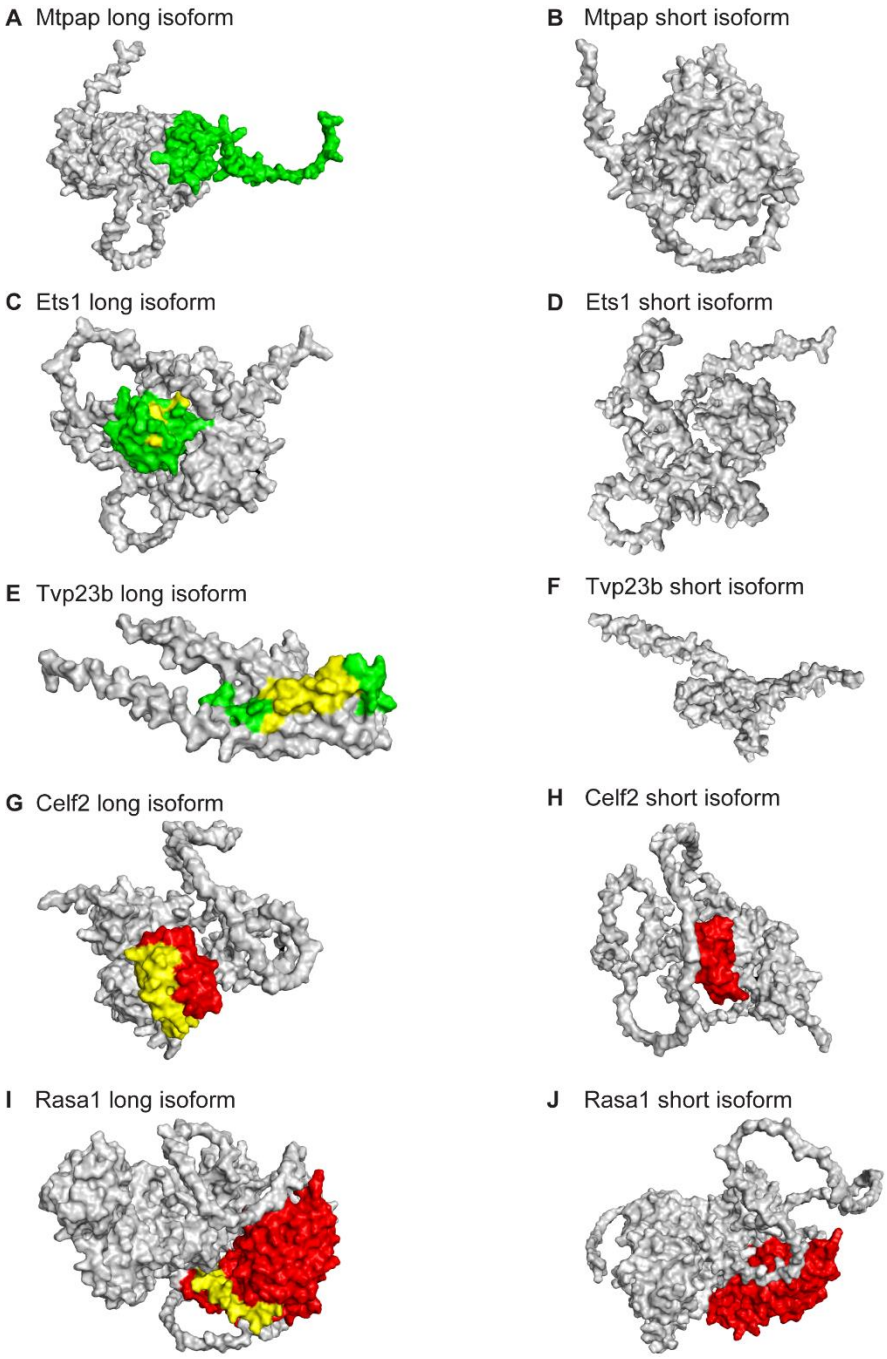


Fig. S16. Predicted protein structures of biomarkers show how AS can influence their structure. (**A**-**J**) AlphaFold2 predicted structures of long (A, C, E, G, I) and short (B, D, F, H, J) AS isoforms. Protein regions retained in long isoforms and important for specified functions were marked green and red respectively, with any overlap marked yellow, and remainder marked grey. AAs 447-585 (green, A) which include AAs 478 and 490, point mutations in which confers susceptibility to fetal-onset spastic ataxia type IV, are retained in long Mtpap isoform (A) but spliced out in short Mtpap isoform (no green,

B). AAs 68-134 (green + yellow, C), which include an ERK2 MAPK docking site (AAs 114,116,120; yellow, C), are retained in long Ets1 isoform (C) but spliced out in short Ets1 isoform (no yellow or green, D). AAs 111-154 (green + yellow, E) which include one of the transmembrane regions (AAs 126-146; yellow, E) are retained in long Tvp23b isoform (E) but spliced out in short Tvp23b isoform (no yellow or green, F). AAs 104-148 (green + yellow) are retained in long Celf2 isoform. AAs 107-187 (yellow + red, G) mark RNA binding site of Celf2 essential for RNA splicing activity. AAs 107-148 (green + yellow, G) mark partial overlap between AS region (AAs 104-148; green + yellow, G) and RNA binding region (AAs 107-187; yellow + red, G) in Celf2 long isoform (G). Splicing out this AS region in short Celf2 isoform (no green or yellow, H) leads to its RNA binding site truncation (red, H). AAs 705-1036 in long Rasa1 isoform (red + yellow, I) mark Ras GTPase activating domain. AAs 967-1011 (yellow, I) within this domain are spliced out in short isoform resulting in truncated GTPase domain (AAs 705-994; red, J).

Fig. S17

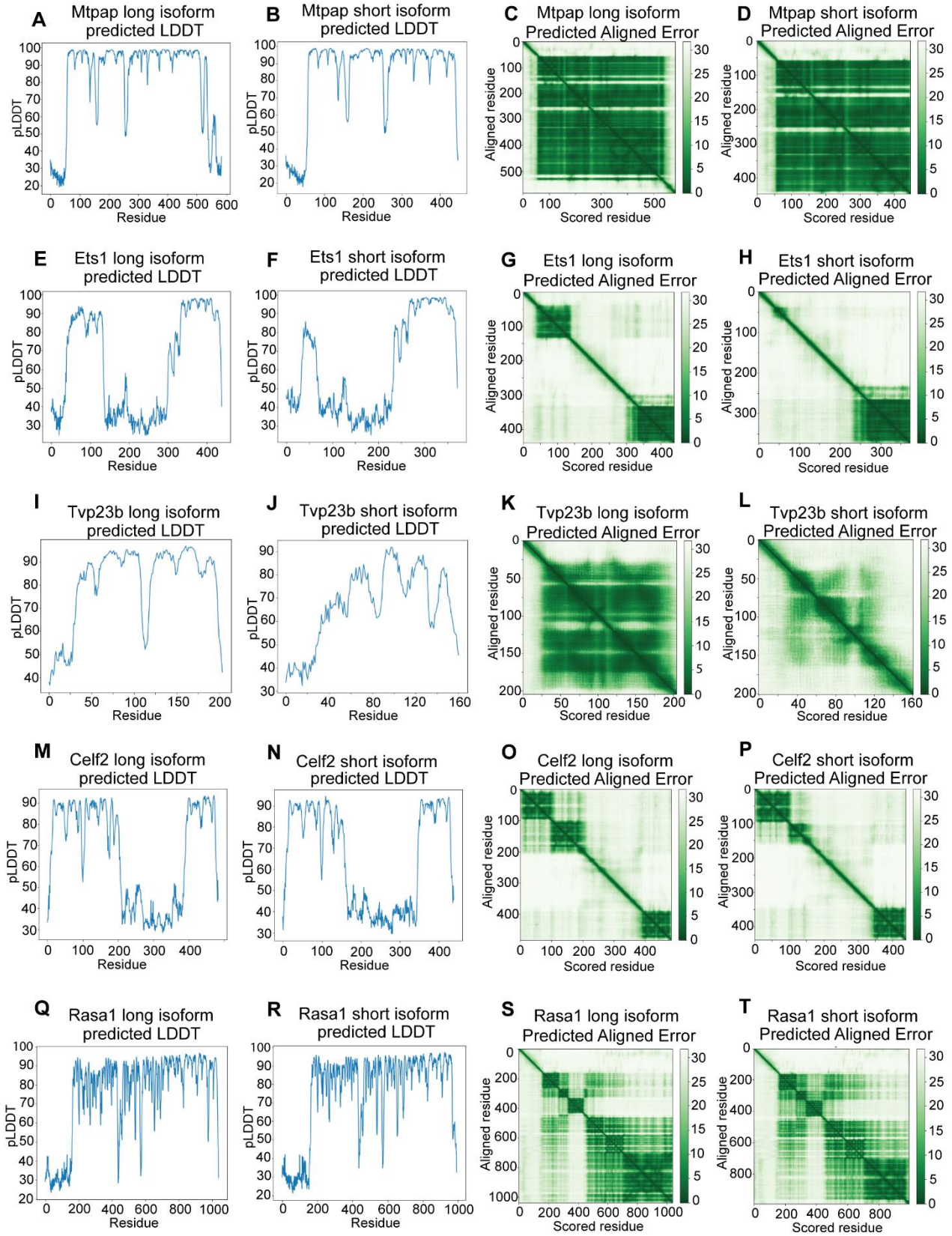
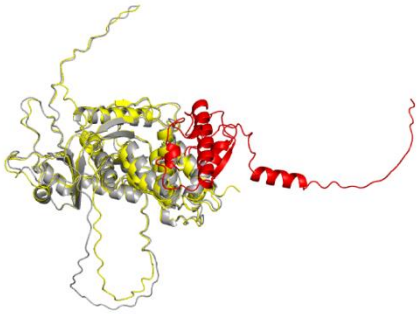


Fig. S17. Confidence plots for AlphaFold2 predicted structures of long and short isoforms of selected biomarkers of interest. (**A**-**T**) Predicted Local Distance Difference

Test (pLDDT) plots (A, B, E, F, I, J, M, N, Q, R) and Predicted Aligned Error plots (C, D, G, H, K, L, O, P, S, T) of AlphaFold2 predicted protein structures for long (A, C, E, G, I, K, M, O, Q, S) and short (B, D, F, H, J, L, N, P, R, T) isoforms of Mtpap (A-D), Ets1 (E-H), Tvp23b (I-L), Celf2 (M-P) and Rasa1 (Q-T). pLDDT is a per-residue measure of AlphaFold2's local confidence, on a scale from 0-100, in the accuracy of each AA residue in its predicted structure. pLDDT plots are a plot of pLDDT scores (y-axis) for each AA residue (x-axis) in the protein structure. Regions with pLDDT scores > 90, 90-70, and 70-50 are expected to be modeled with high, medium, and low accuracy respectively. Regions with pLDDT scores < 50 is a strong predictor for disorder, i.e those regions are either intrinsically disordered or only structured as part of a complex. Predicted Aligned Error (x-y) plots report AlphaFold2's expected position error at residue x (x-axis), when the predicted and true structures are aligned on residue y (y-axis). Predicted Aligned Error is useful for assessing confidence in relative domain positions (i.e., domain packing). For residues x and y drawn from two different domains, a consistently low Predicted Aligned Error at (x, y) suggests AlphaFold2 algorithm is confident about the relative domain positions, and vice versa.

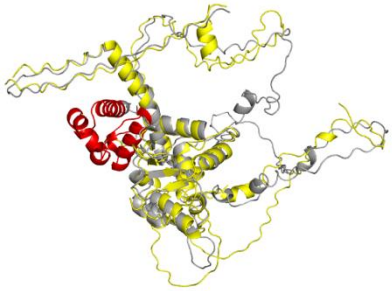
Fig. S18

A Superimposed Mtpap long & short isoform with 1 twist



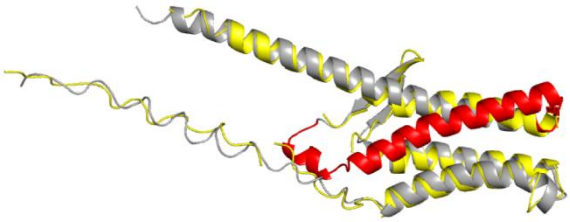
Mtpap long isoform in gray
Mtpap short isoform in yellow
Mtpap short isoform spliced out region in red
RMSD of equivalent positions = 2.85Å

B Superimposed Ets1 long & short isoform with 4 twists



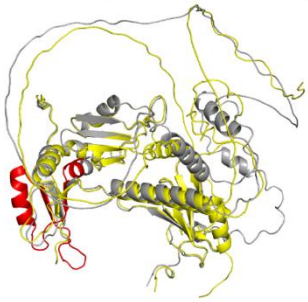
Ets1 long isoform in gray
Ets1 short isoform in yellow
Ets1 short isoform spliced out region in red
RMSD of equivalent positions = 4.03Å

C Superimposed Tvp23b long & short isoform with 4 twists



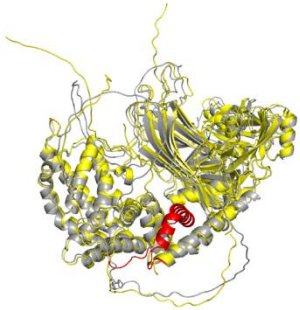
Tvp23b long isoform in gray
Tvp23b short isoform in yellow
Tvp23b short isoform spliced out region in red
RMSD of equivalent positions = 4.24Å

D Superimposed Celf2 long & short isoform with 4 twists



Celf2 long isoform in gray
Celf2 short isoform in yellow
Celf2 short isoform spliced out region in red
RMSD of equivalent positions = 3.60Å

E Superimposed Rasa1 long & short isoform with 5 twists



Rasa1 long isoform in gray
Rasa1 short isoform in yellow
Rasa1 short isoform spliced out region in red
RMSD of equivalent positions = 2.77Å

Fig. S18. Superimposition of predicted structures of long and short biomarker isoforms show preservation of global structures of short isoforms in spite of loss

of their spliced out regions. (**A-E**) Superimposition of AlphaFold2 predicted structures of long (grey) and short (yellow) AS isoforms of Mtpap (A), Ets1 (B), Tvp23b (C), Celf2 (D), and Rasa1 (E) via FATCAT algorithm. The protein regions retained in long isoforms were marked red. RMSDs of aligned Cα atoms between long and short isoforms are indicated.

Fig. S19

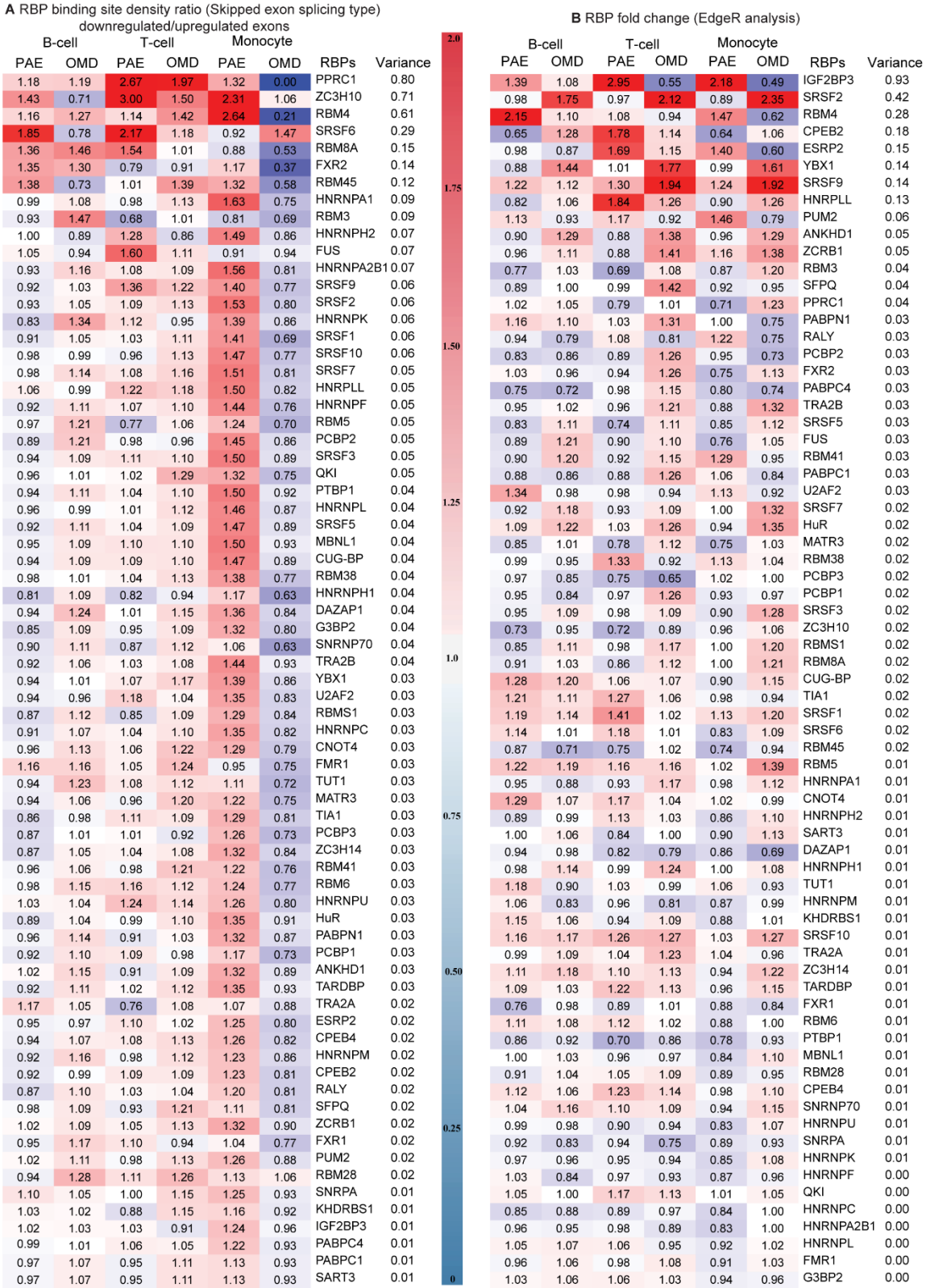


Fig. S19. Heatmap of binding site density ratio of RBPs and fold-change in RBP expression in peripheral immune cells of PAE and OMD mice. (**A**) Heatmap show RBP binding site density ratio, as determined by the ratio of RBP binding site density in

downregulated exons (negative Δpsi splicing events) vs upregulated exons (positive Δpsi splicing events) in skipped exon AS type; in B-cells, T-cells, and monocytes; in PAE and OMD mice. (**B**) Heatmap show differential RBP expression, as determined by EdgeR analysis; in B-cells, T-cells, and monocytes; in PAE and OMD mice. (A ,B) The RBPs are arranged in descending order of their variance among different experimental conditions. The color range of heatmap cells is indicated in the middle between panels a and b.

Fig. S20

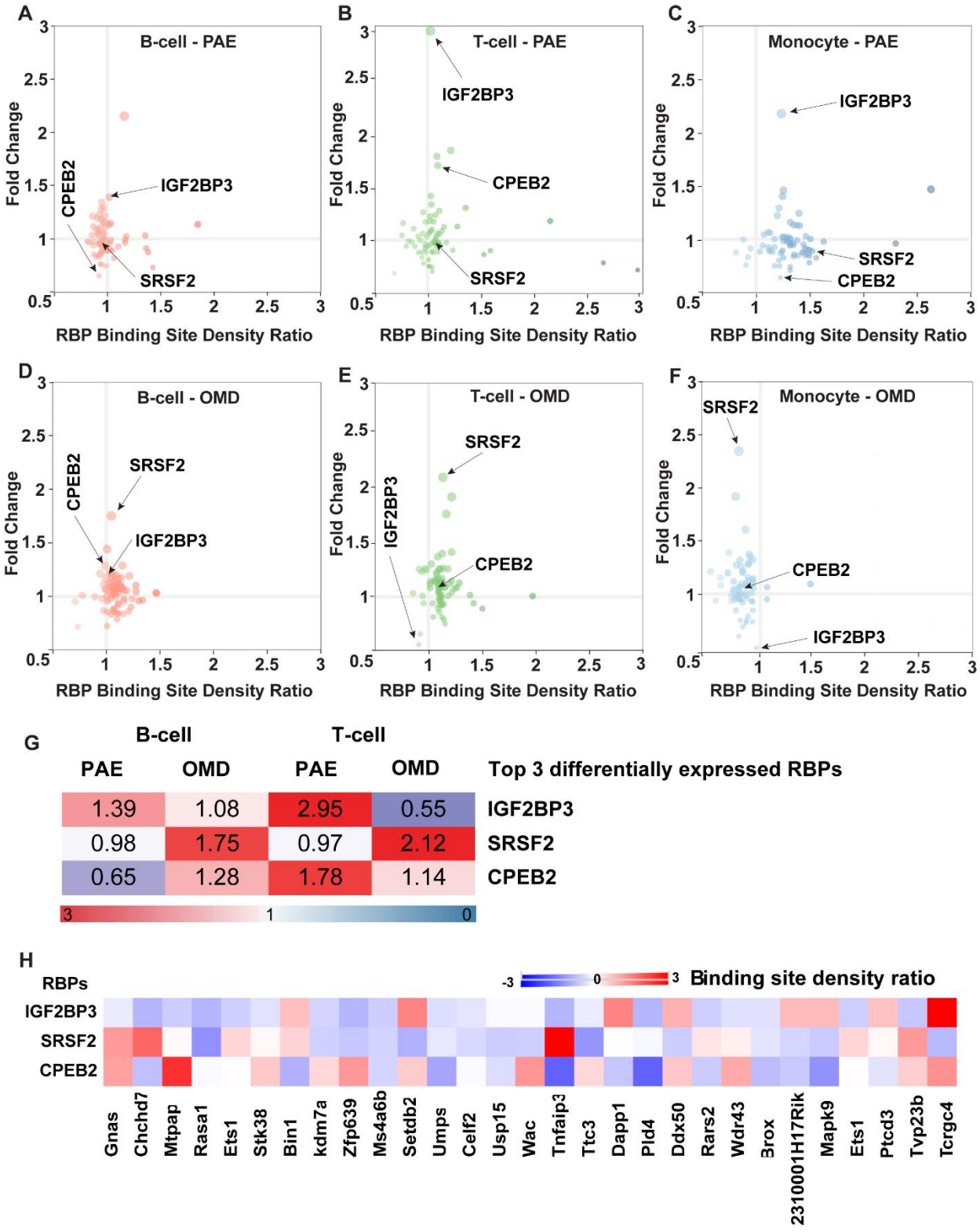


Fig. S20. Differential expression of RBPs may underlie opposing AS pattern observed in PBMCs of PAE and OMD. (**A-F**) Scatter plots of RBP binding site density ratio of RBPs in SE AS type versus differential expression of those RBPs in B-cells (A, D),

T-cells (B, E), and monocytes (C, F) in PAE (A-C) and OMD (D-F). RBP binding site density ratio is determined as ratio of RBP binding site density (number of RBP binding sites per million base pairs) in downregulated (-Δpsi) vs upregulated (+Δpsi) exons in SE AS type. IGF2BP3, CPEB2, and SRSF2; among the highest differentially expressed RBPs between PAE and OMD across all cell-types; are annotated. (**G**) Heatmap shows fold change in their expression in B-cells and T-cells in PAE and OMD. (**H**) Heatmap shows their binding site density in biomarkers, around their splice relevant sites.

Table 1

	OMD-control			OMD			p		
	All	Male	Female	All	Male	Female	All	Male	Female
n	30	13	17	26	16	10			
Body weight (g)	23.8±3.7	26.7±3.7	21.6±2.0	21.2±3.2	22.1±3.4	19.7±2.4	0.029*	0.010*	>0.999
Fasting blood glucose (mg/dl)	135.6±22.5	153.8±17.3	121.6±14.4	123.6±16.7	127.6±19.2	117.4±9.5	0.508	0.036*	>0.999
Random blood glucose (mg/dl)	166.3±21.4	185.1±11.3	151.9±15.0	163.7±21.1	171.2±19.9	151.7±18.0	0.996	0.376	>0.999

Table S1. **Metabolic assessment of OMD and OMD-control mice at postnatal day 29.** The offspring of the MD group (OMD group) shows lower body weight and blood glucose level at 6 hours post fasting at postnatal day 29 (P29) in males, but not females, compared to the offspring of the MD control group (OMD-control group). Statistical significance was assessed by Kruskal-Wallis test (Body weight and Fasting blood glucose) or one-way ANOVA (Random blood glucose).

Movie S1. Predicted protein structures of the long and short isoforms of Mtpap, Ets1, Tvp23b, Celf2 and Rasa1. The AlphaFold2 predicted protein structures of the long and short isoforms of Mtpap, Ets1, Tvp23b, Celf2 and Rasa1 respectively, as annotated in Fig. 8, and their FATCAT superimposed structures as annotated in Fig. 9, displayed in tandem.

Supplementary References:

1. M. M. Buitrago, J. B. Schulz, J. Dichgans, A. R. Luft, Short and long-term motor skill learning in an accelerated rotarod training paradigm. *Neurobiol. Learn. Mem.* **81**, 211–216 (2004).

2. Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data (July 27, 2021).

3. D. Kim, J. M. Paggi, C. Park, C. Bennett, S. L. Salzberg, Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915 (2019).

4. A. Dobin, *et al.*, STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

5. S. Anders, P. T. Pyl, W. Huber, HTSeq — a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).

6. Y. Liao, G. K. Smyth, W. Shi, featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).

7. V. Jiménez-Jacinto, A. Sanchez-Flores, L. Vega-Alvarado, Integrative Differential Expression Analysis for Multiple EXperiments (IDEAMEX): A Web Server Tool for Integrated RNA-Seq Data Analysis. *Front. Genet.* **10**, 279 (2019).

8. S. Shen, *et al.*, rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proc Natl Acad Sci USA* **111**, E5593-601 (2014).

9. Y. I. Li, *et al.*, Annotation-free quantification of RNA splicing using LeafCutter. *Nat. Genet.* **50**, 151–158 (2018).

10. M. V. Kuleshov, *et al.*, Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44**, W90-7 (2016).

11. D. Barrell, *et al.*, The GOA database in 2009--an integrated Gene Ontology Annotation resource. *Nucleic Acids Res.* **37**, D396-403 (2009).

12. F. Supek, M. Bošnjak, N. Škunca, T. Šmuc, REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS ONE* **6**, e21800 (2011).

13. MDSJ - Multidimensional Scaling for Java (June 4, 2021).

14. Tableau · GitHub (June 4, 2021).

15. deepmind/alphafold: Open source code for AlphaFold. (July 27, 2021).

16. J. Jumper, *et al.*, Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

17. A. Porollo, J. Meller, Versatile annotation and publication quality visualization of protein complexes using POLYVIEW-3D. *BMC Bioinformatics* **8**, 316 (2007).

18. Z. Li, L. Jaroszewski, M. Iyer, M. Sedova, A. Godzik, FATCAT 2.0: towards a better understanding of the structural diversity of proteins. *Nucleic Acids Res.* **48**, W60–W64 (2020).

19. J. Y. Hwang, *et al.*, rMAPS2: an update of the RNA map analysis and plotting server for alternative splicing regulation. *Nucleic Acids Res.* **48**, W300–W306 (2020).

20. I. Paz, I. Kosti, M. Ares, M. Cline, Y. Mandel-Gutfreund, RBPmap: a web server for mapping binding sites of RNA-binding proteins. *Nucleic Acids Res.* **42**, W361-7 (2014).

21. linuxpham/biovinci (July 28, 2021).

22. Z. Gu, L. Gu, R. Eils, M. Schlesner, B. Brors, circlize Implements and enhances circular visualization in R. *Bioinformatics* **30**, 2811–2812 (2014).

23. J. D. Hunter, Matplotlib: A 2D Graphics Environment. *Comput. Sci. Eng.* **9**, 90–95 (2007).

24. M. Waskom, *et al.*, Seaborn: V0.5.0 (November 2014). *Zenodo* (2014) https:/doi.org/10.5281/zenodo.12710.

25. morpheus.js/README.md at master · cmap/morpheus.js (July 28, 2021).