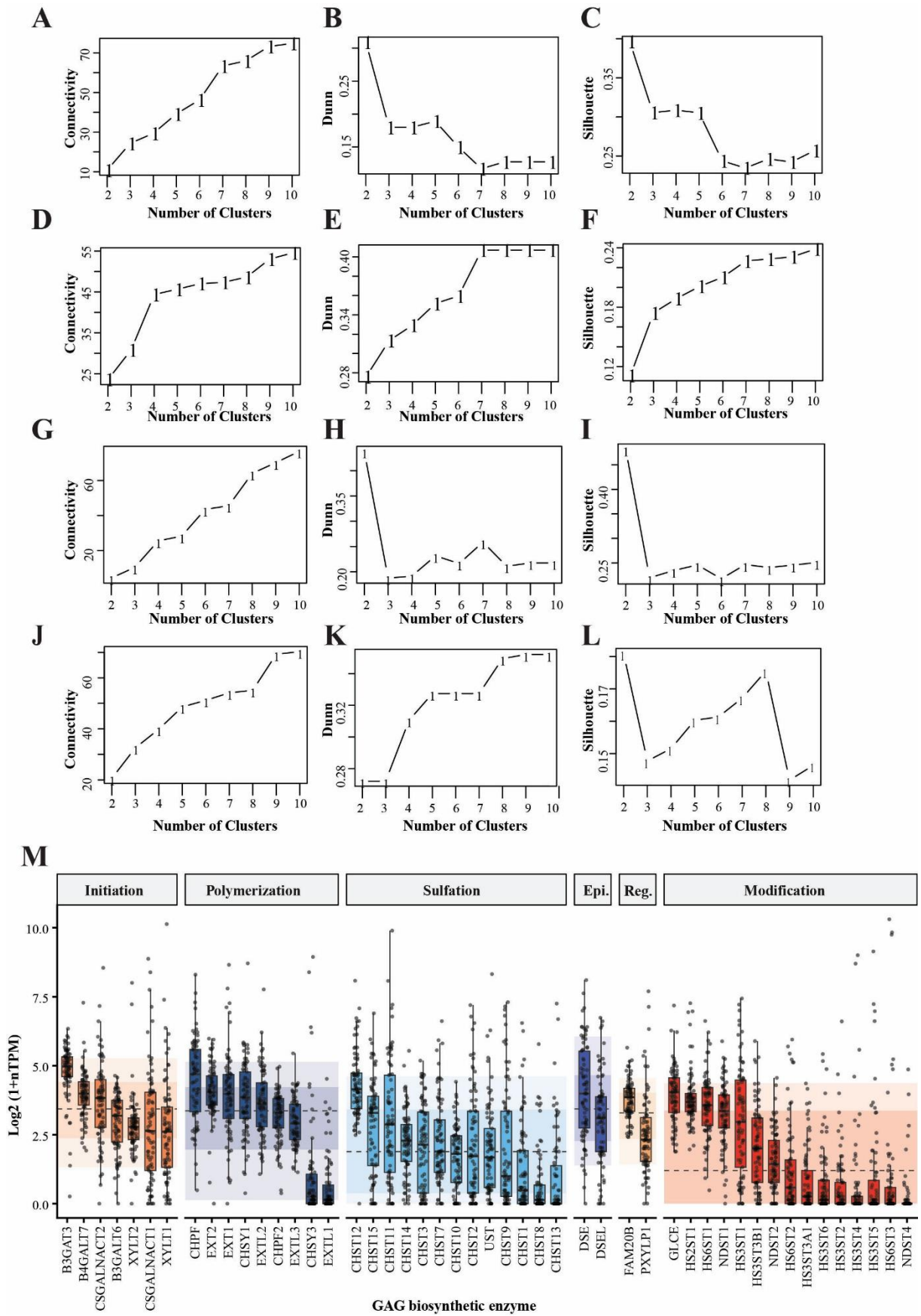
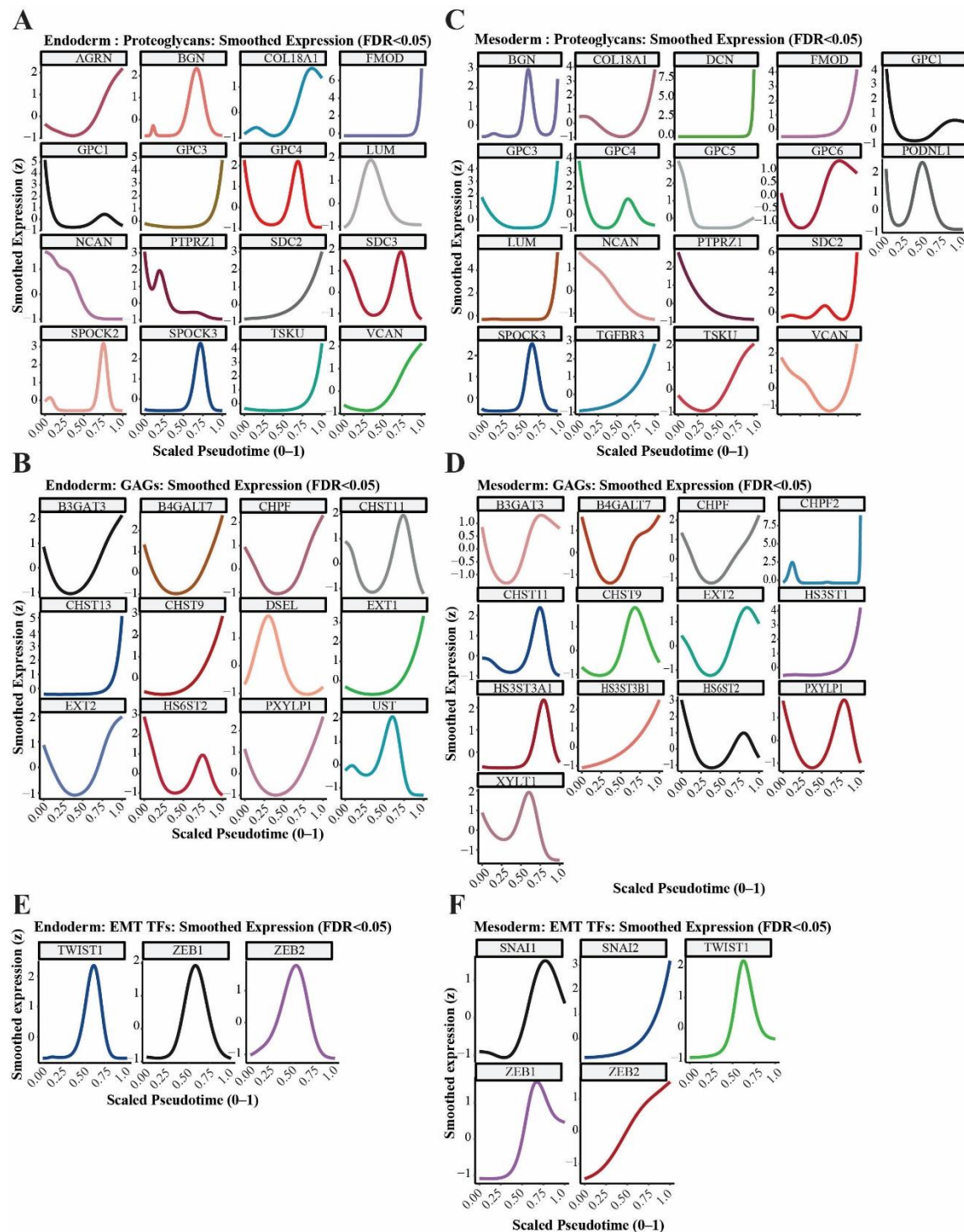


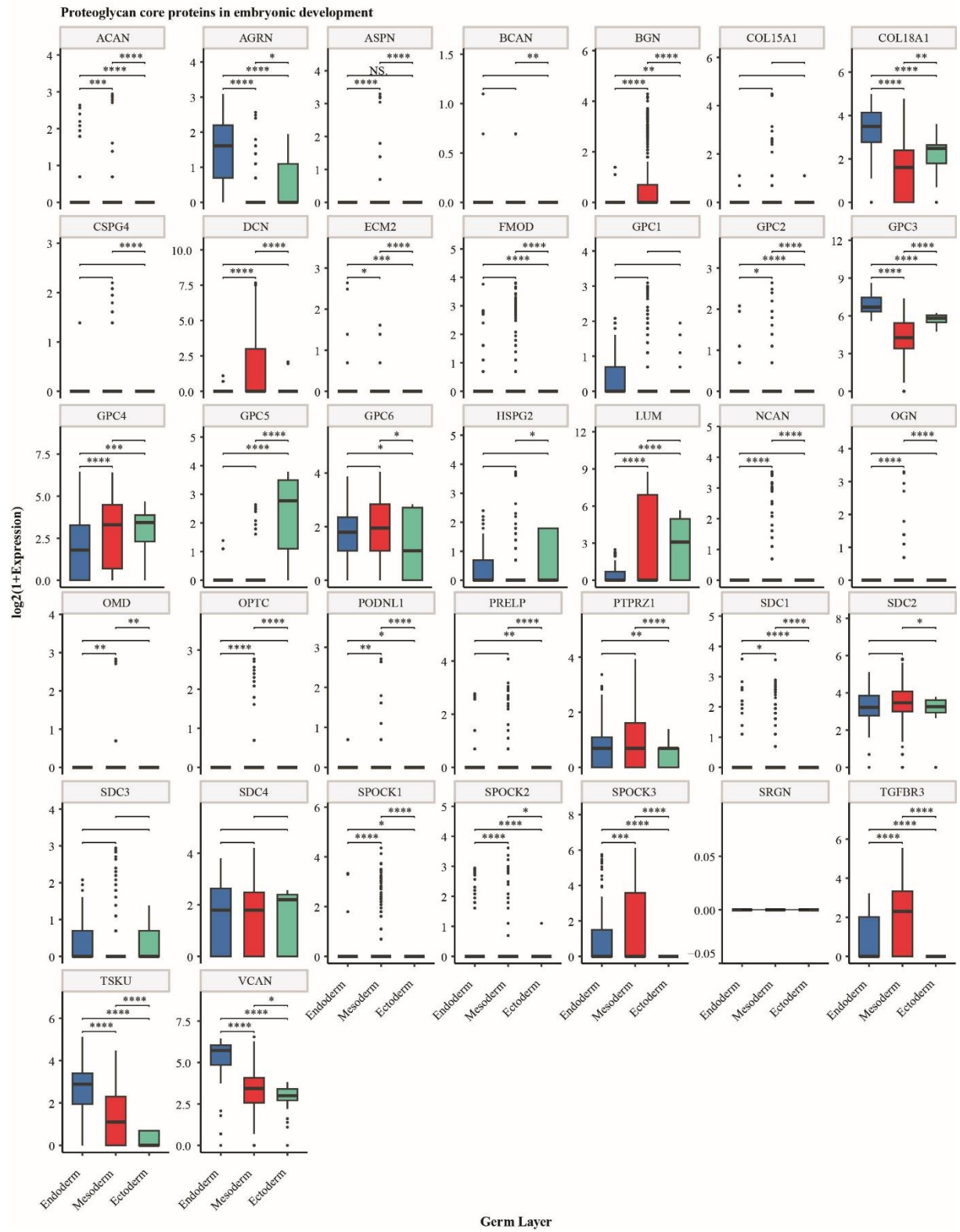


Supplementary Figure 1: Cluster stability for proteoglycan and GAG hierarchical clustering. (A-C) Stability metrics (Connectivity, Dunn, Silhouette) for cell-type clustering in the proteoglycan heatmap. (D-F) The same metrics for proteoglycan gene clustering. (G-I) Stability metrics for cell-type clustering in the GAG heatmap. (J-

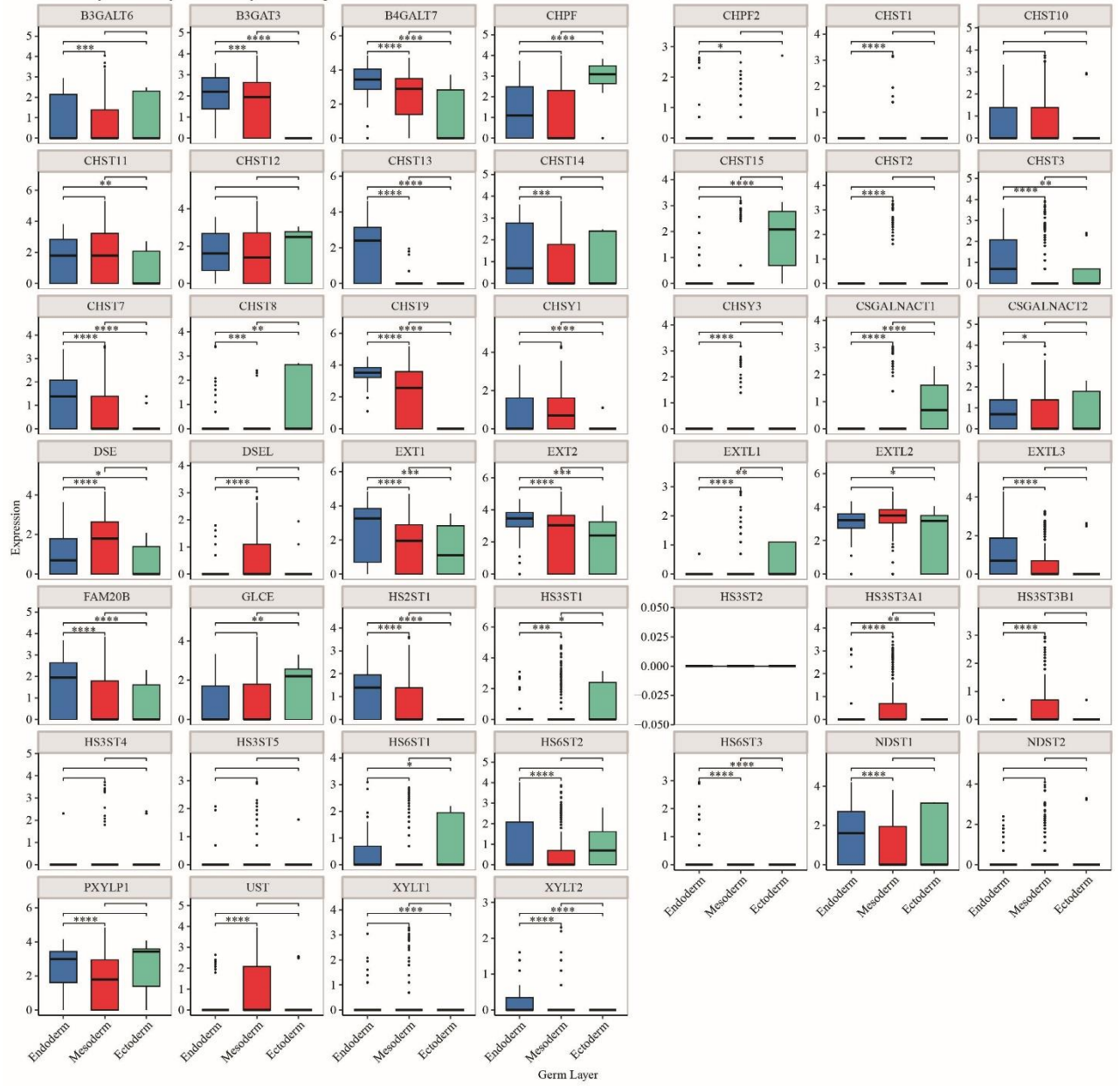
L) Stability metrics for GAG gene clustering. Higher Silhouette and Dunn values and lower Connectivity indicate more stable cluster solutions. Metrics are shown for 2-10 clusters. **(M)** Distribution of GAG biosynthetic enzymes expression across adult cell types. Boxplots show $\log_2(1+nTPM)$ expression of individual GAG biosynthetic enzymes across cell types, with dots representing individual cell types. Genes are grouped according to enzymatic subgroup. Shaded regions indicate the interquartile range of expression values across all enzymes and cell types within each subgroup, and dashed lines indicate subgroup medians.



G

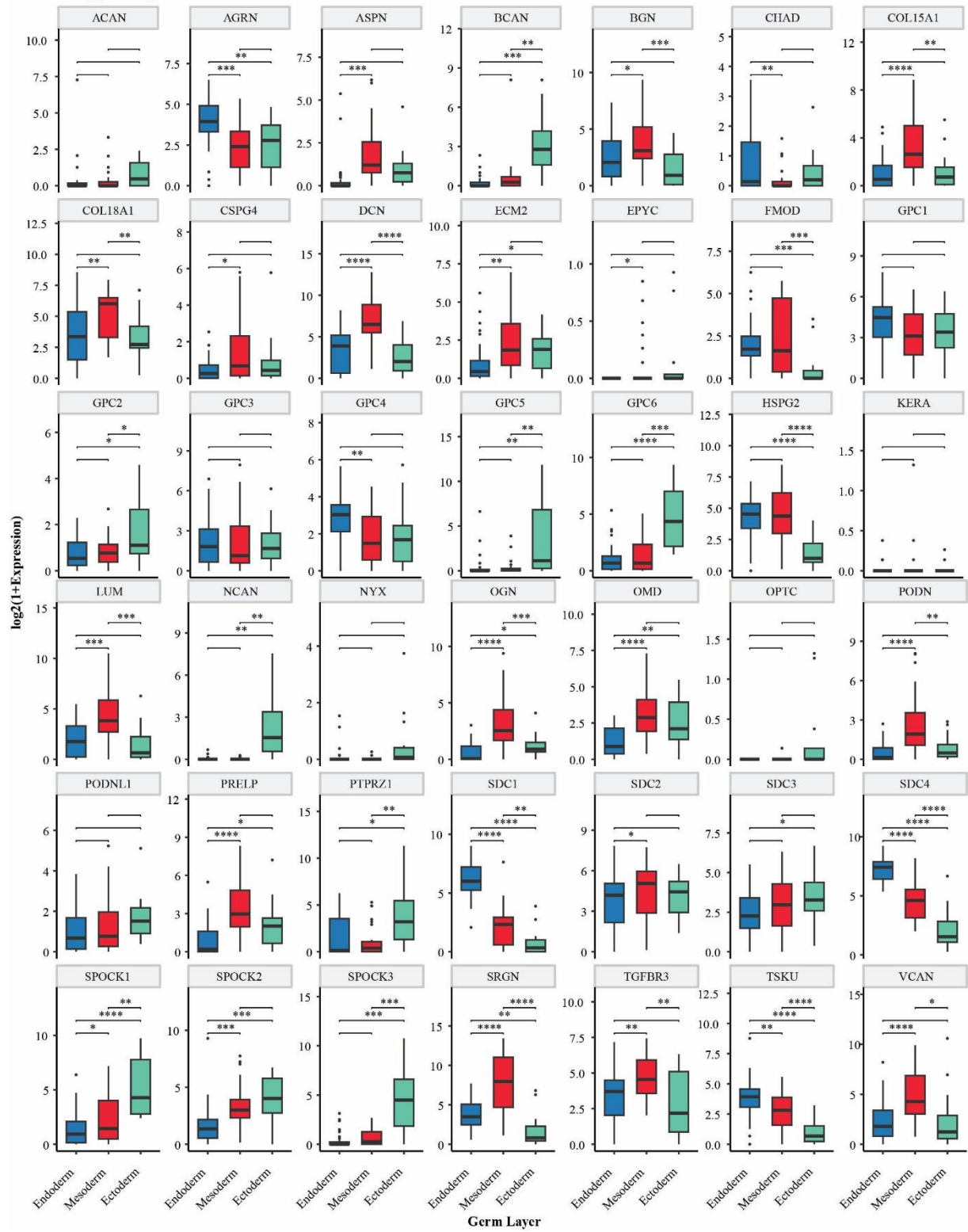


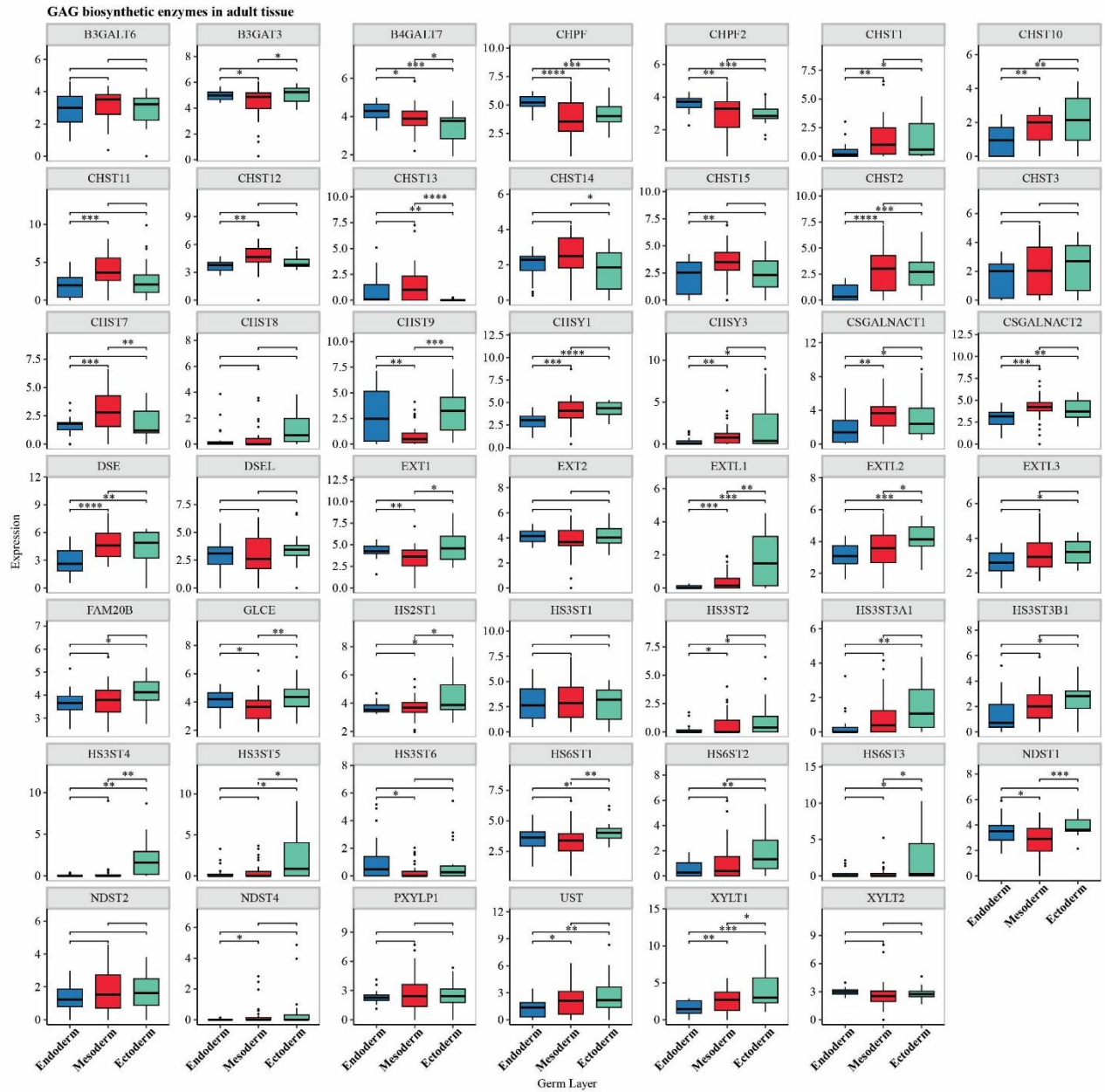
GAG biosynthetic enzymes in embryonic development



H

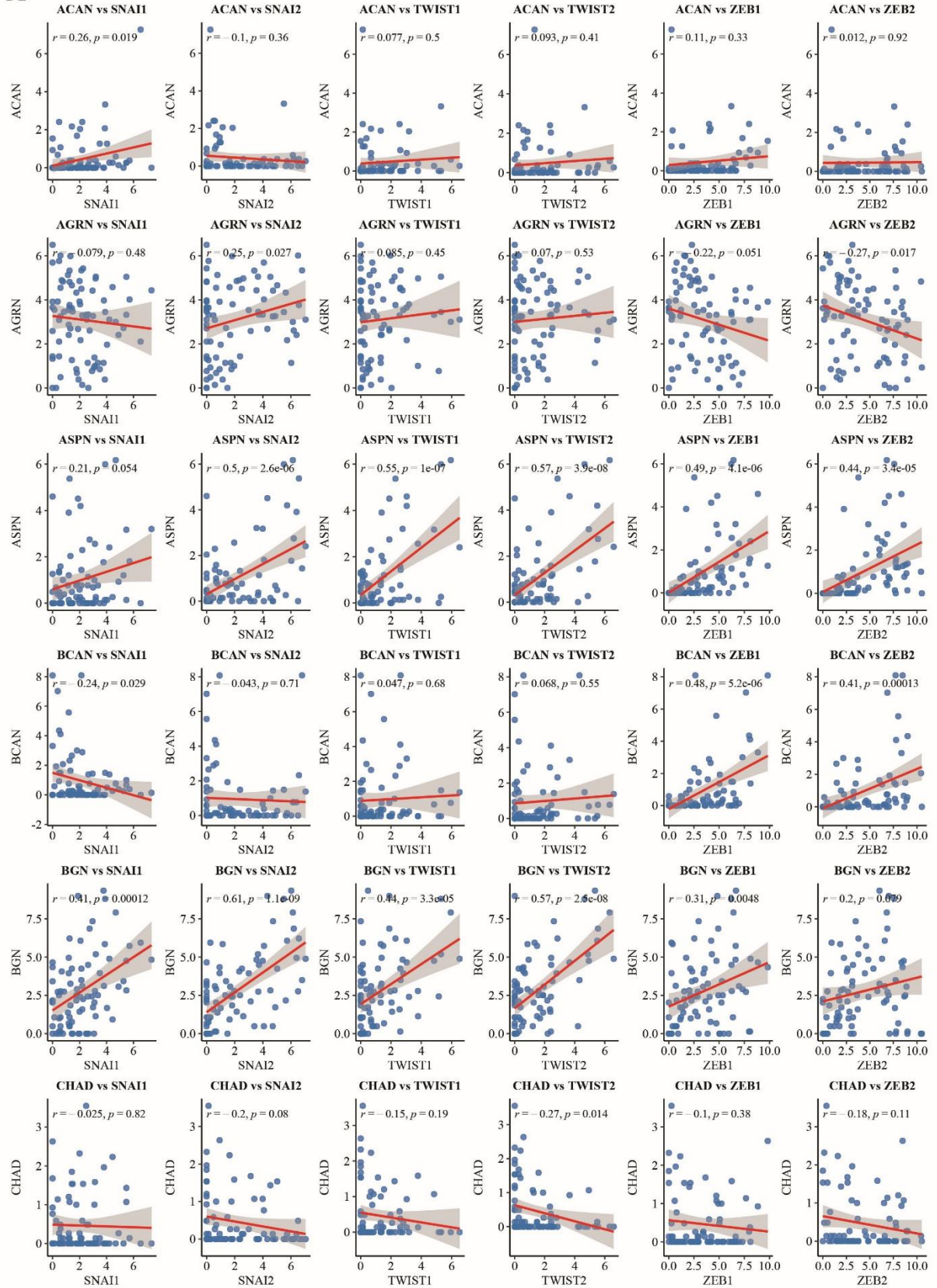
Proteoglycan core proteins in adult tissue

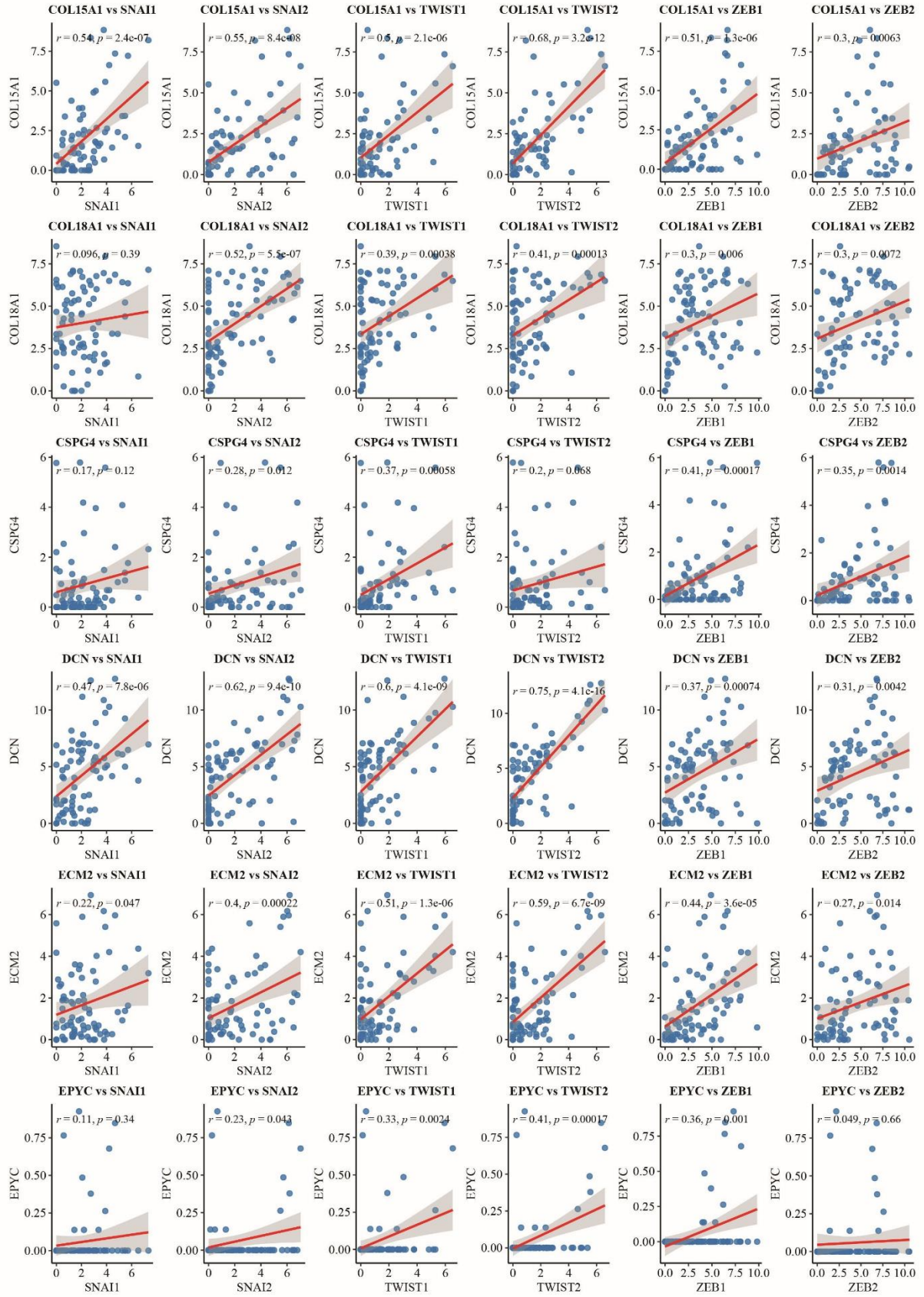


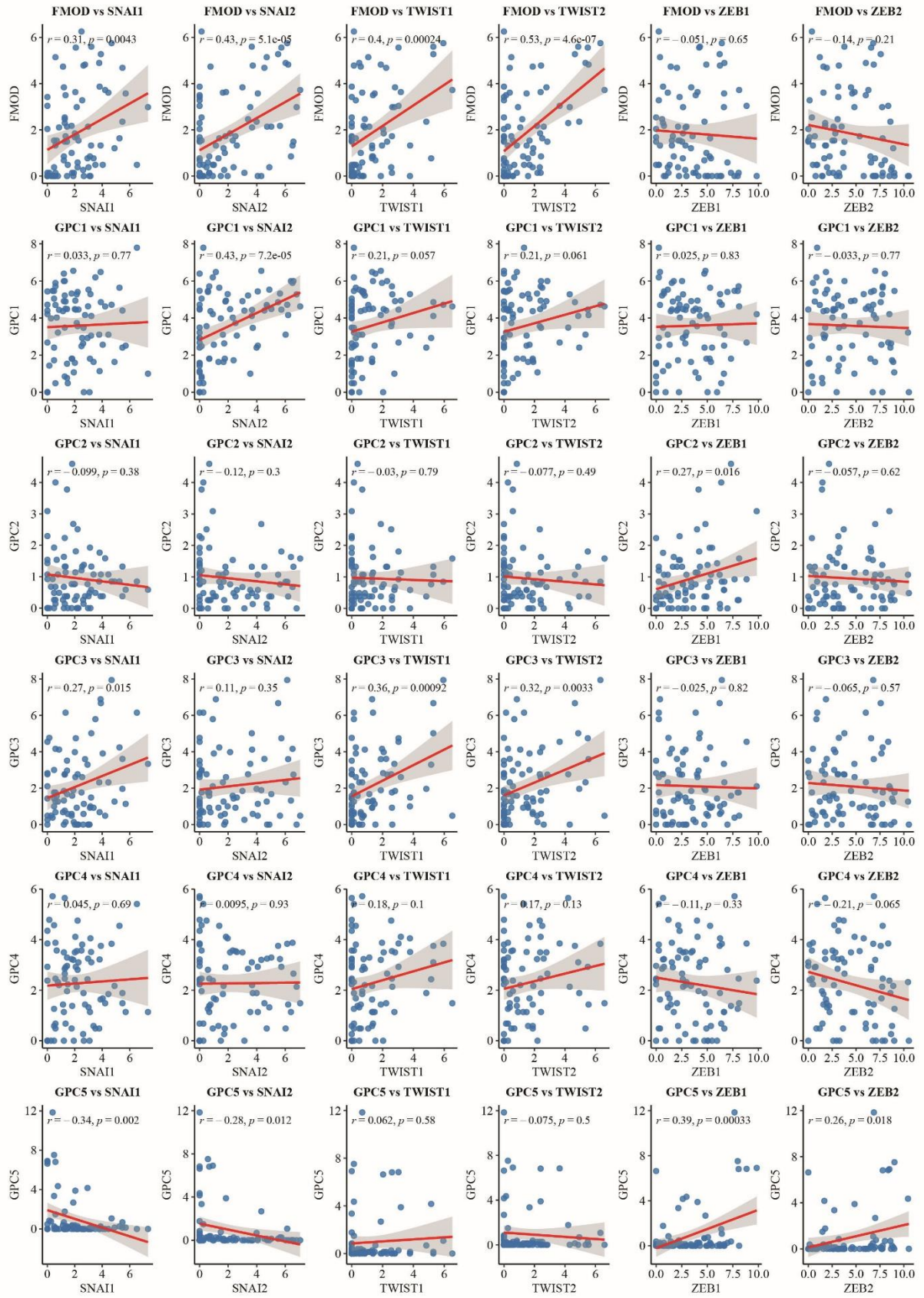


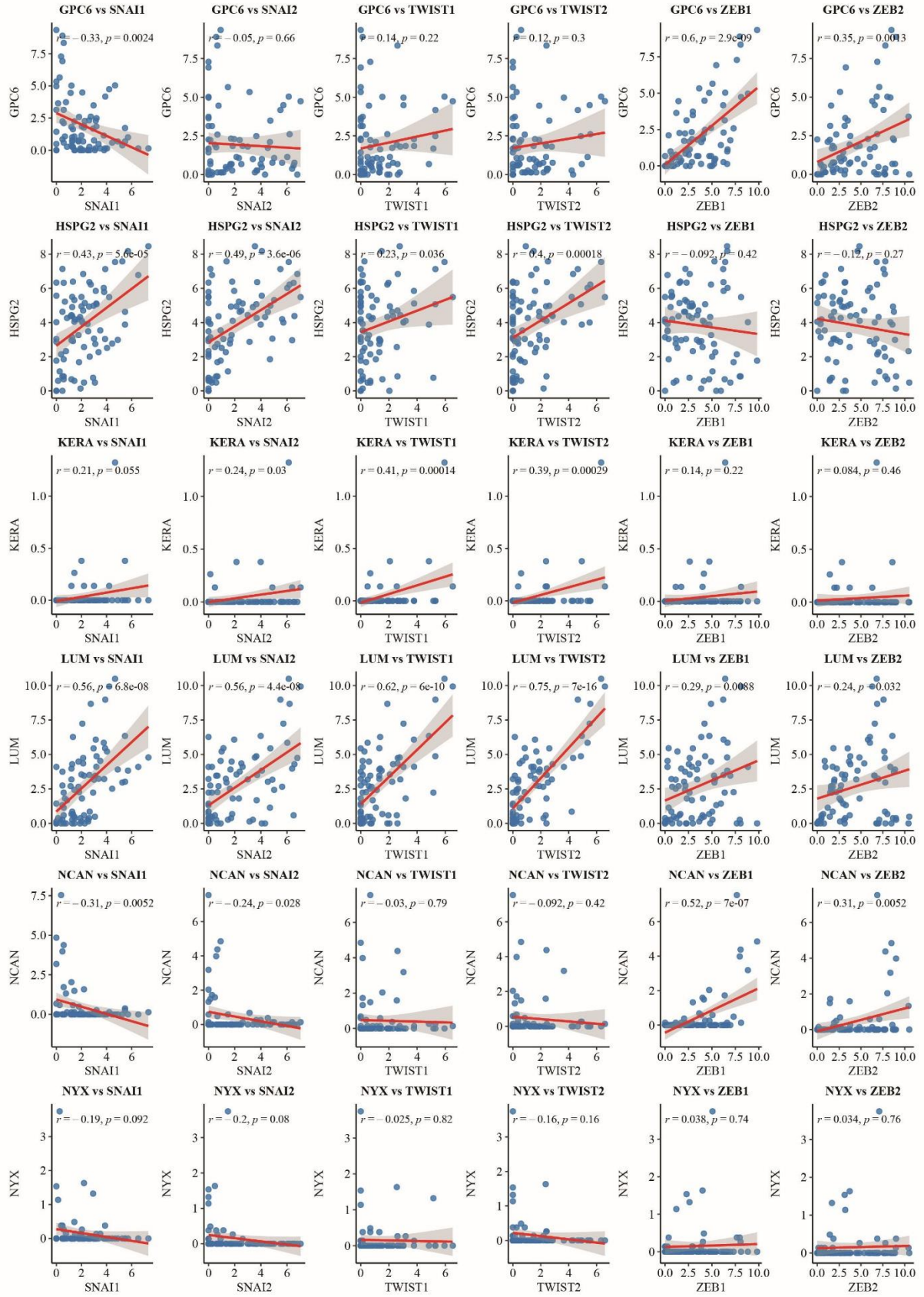
Supplementary Figure 2: Gene-expression dynamics across gastrulation pseudotime and adult tissue origins. (A–D) Smoothed pseudotime expression curves for proteoglycan genes (A, C) and GAG genes (B, D) across the endoderm (A–B) and mesoderm (C–D) trajectories. Curves indicate whether each gene shows early, increasing, decreasing, or fluctuating expression over pseudotime. (E–F) Smoothed expression profiles of EMT transcription factors along the endodermal (E) and mesodermal (F) trajectories. (G) Boxplots showing metacell-level expression of proteoglycan and GAG genes across embryonic germ-layer lineages (endoderm, mesoderm, ectoderm). (H) Boxplots showing expression across adult tissue origins, summarizing proteoglycan and GAG-pathway gene expression in Human Protein Atlas cell types.

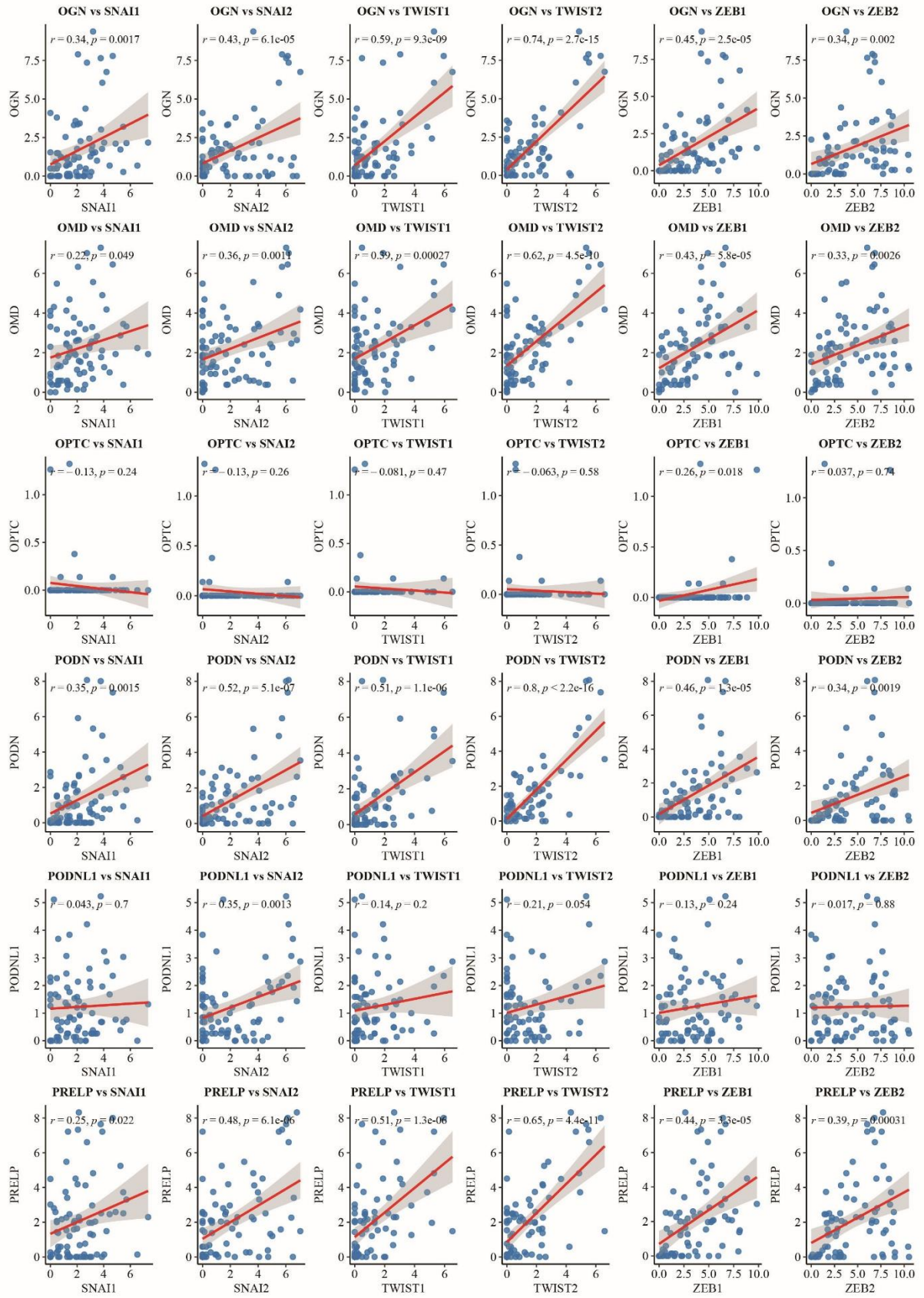
A Correlation between proteoglycan core proteins and EMT TFs

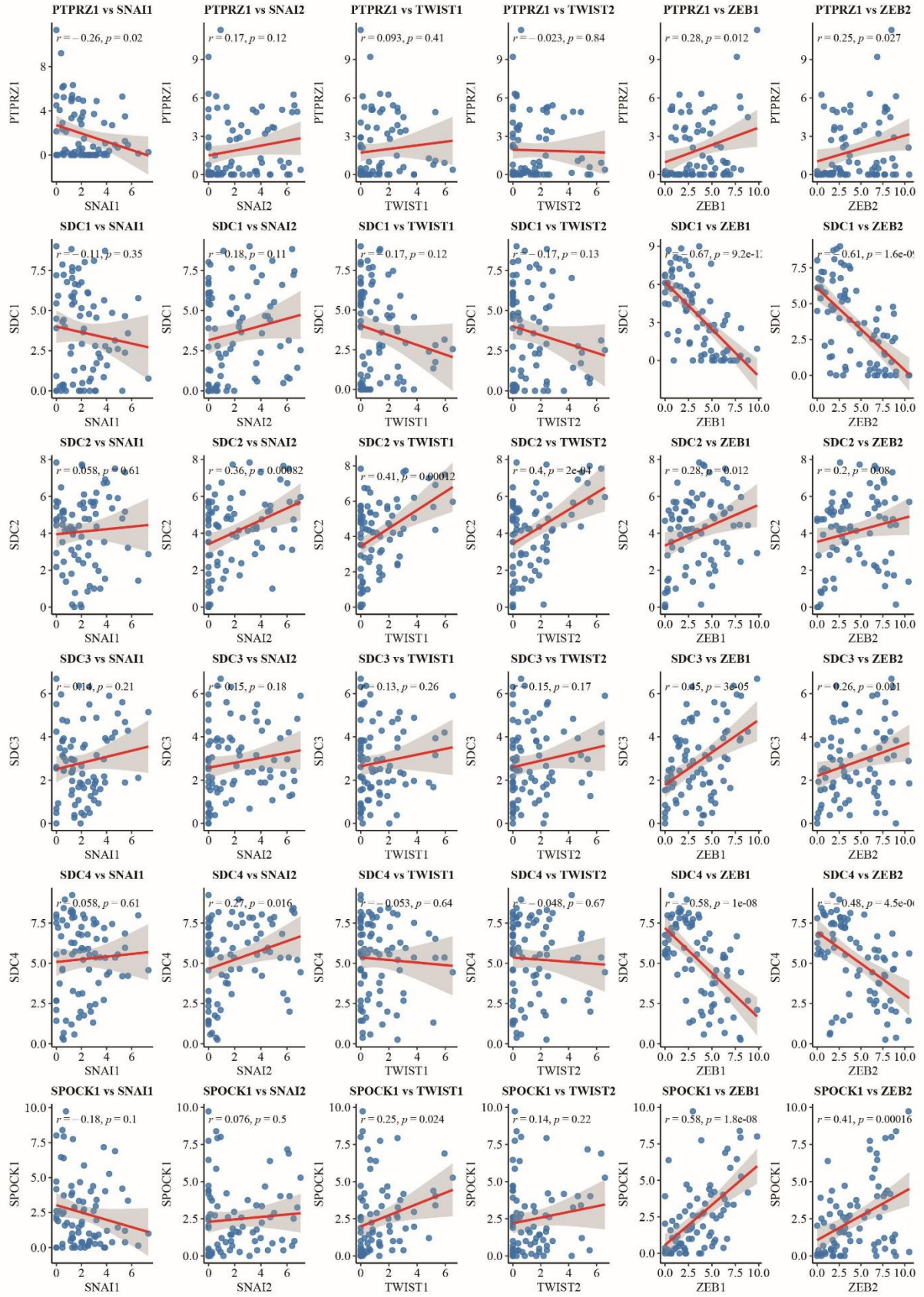


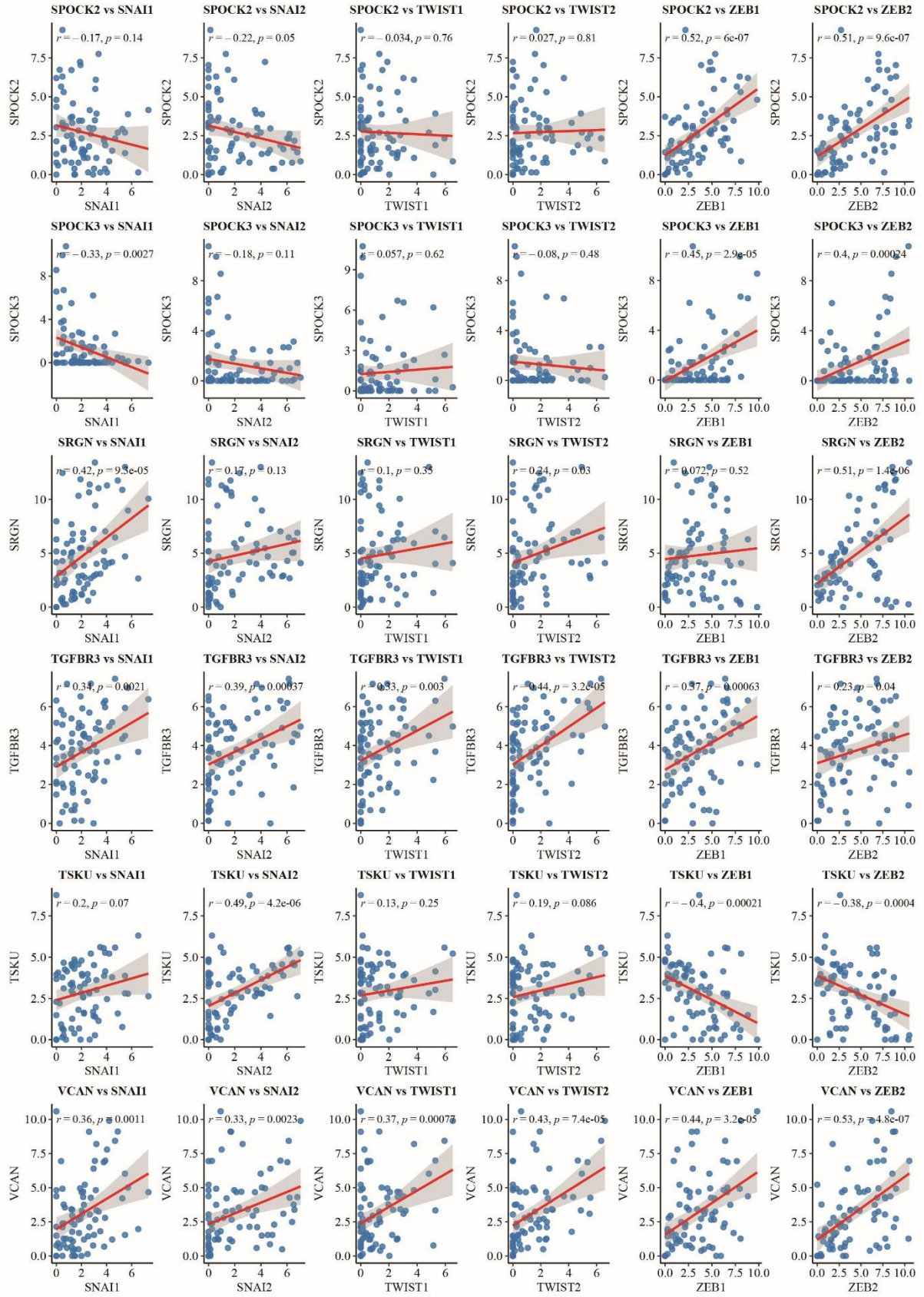


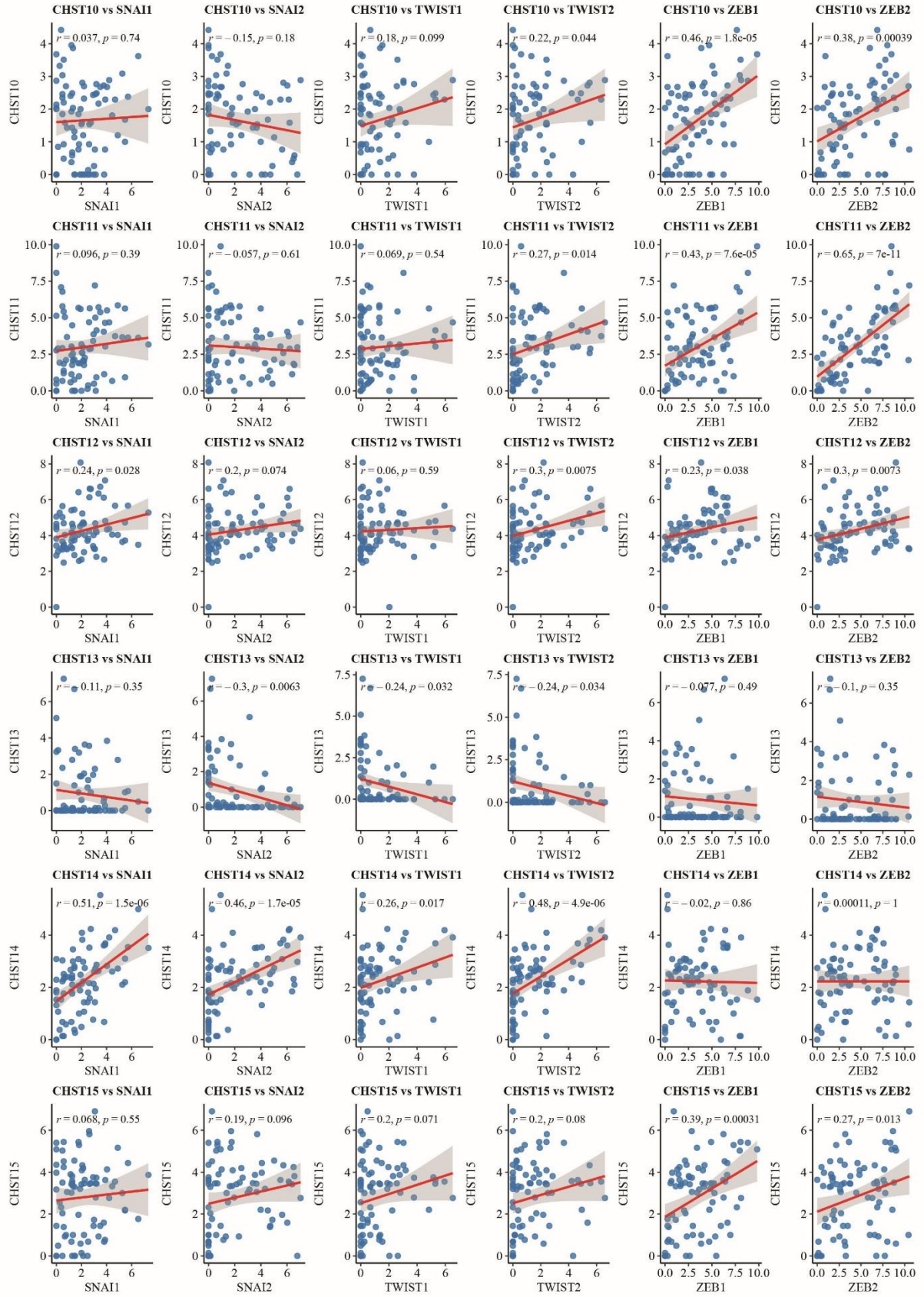


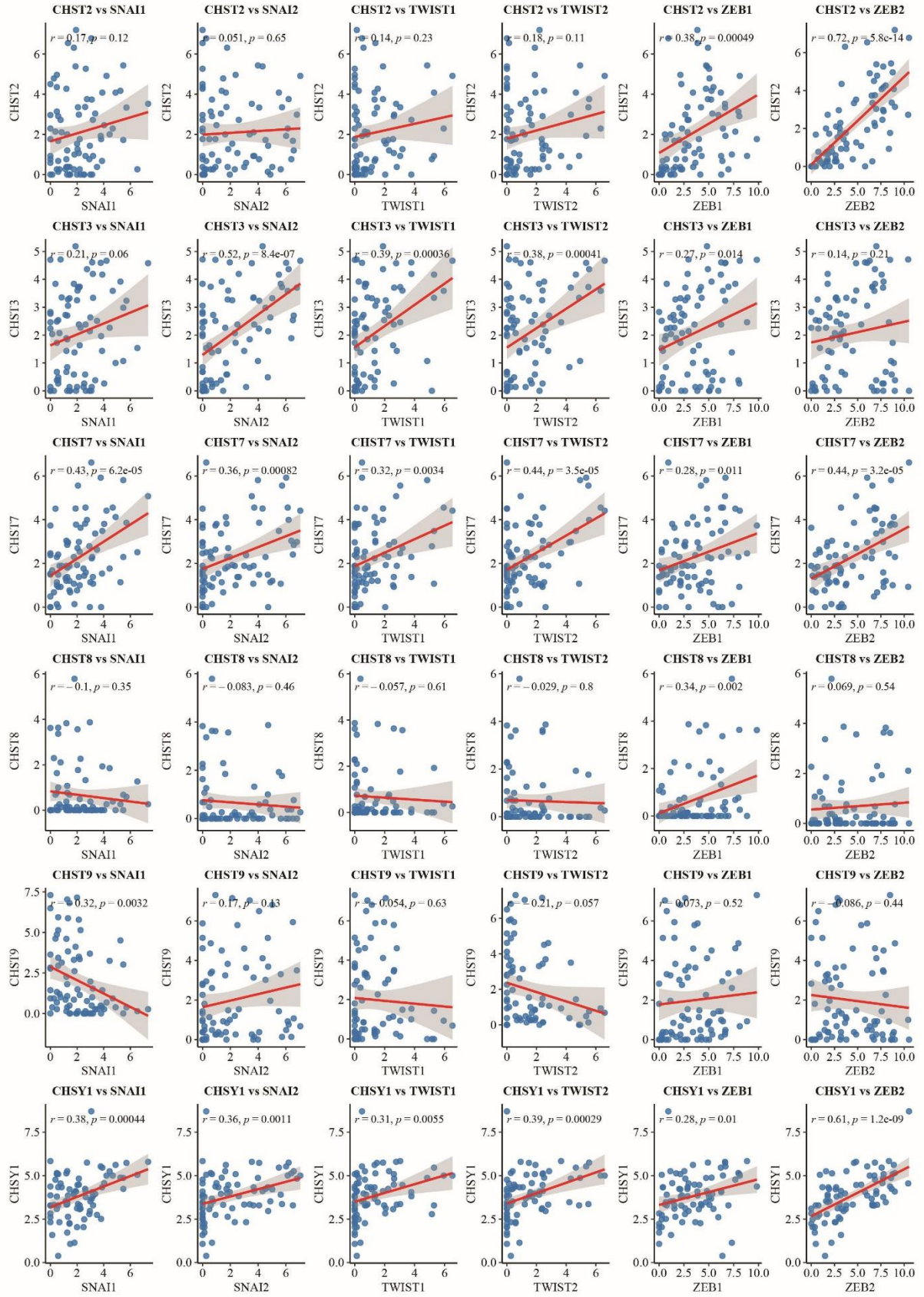


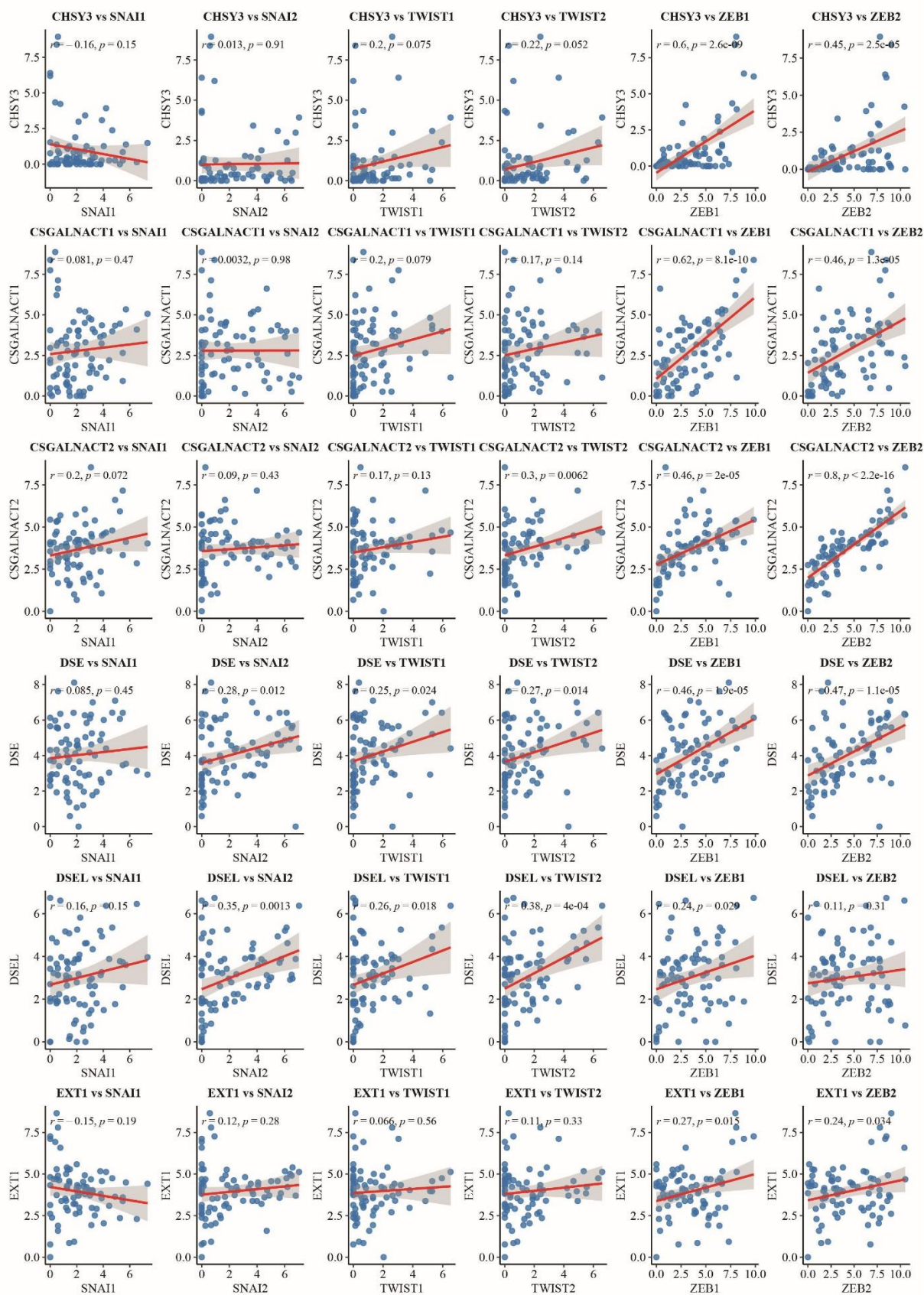


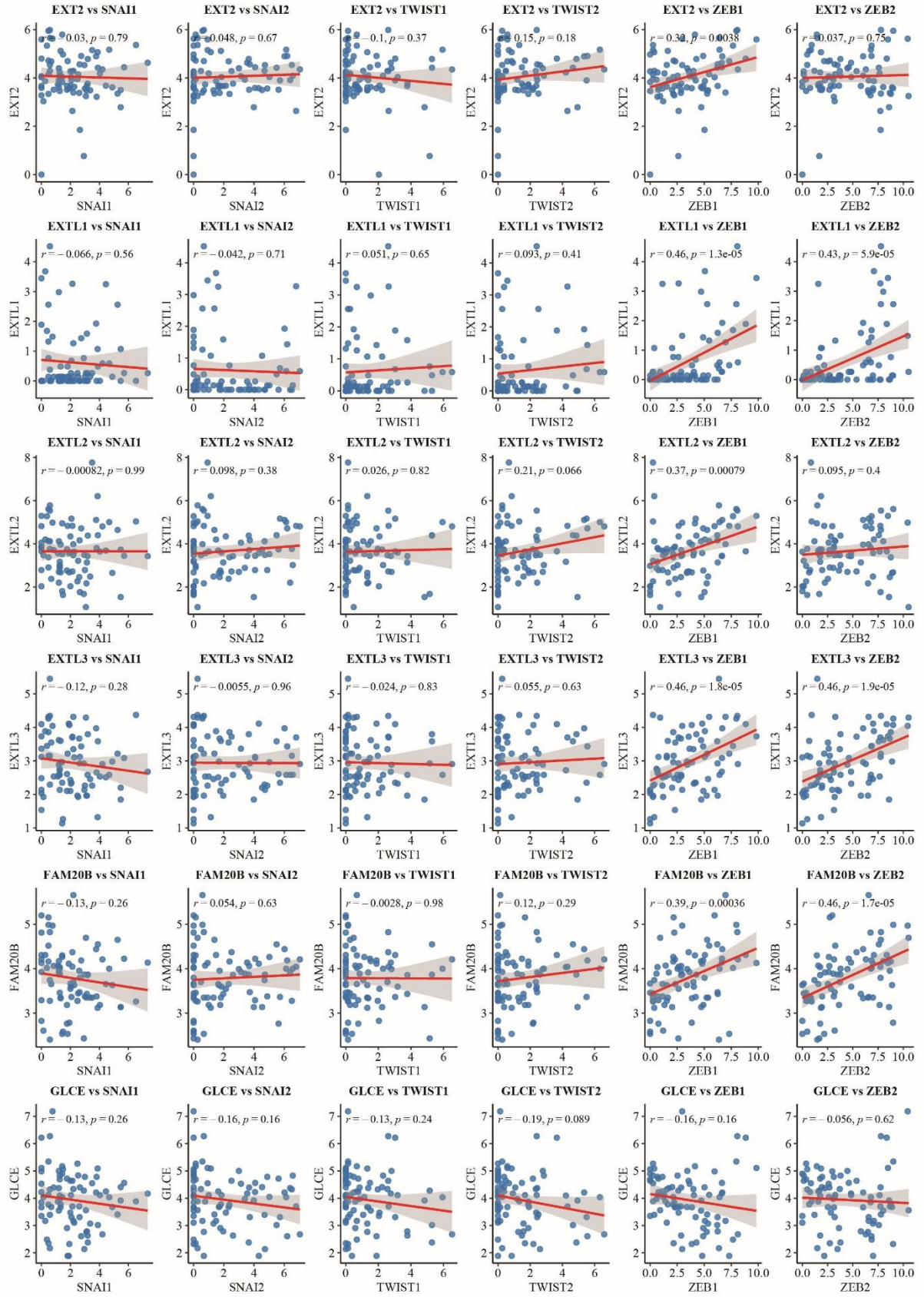


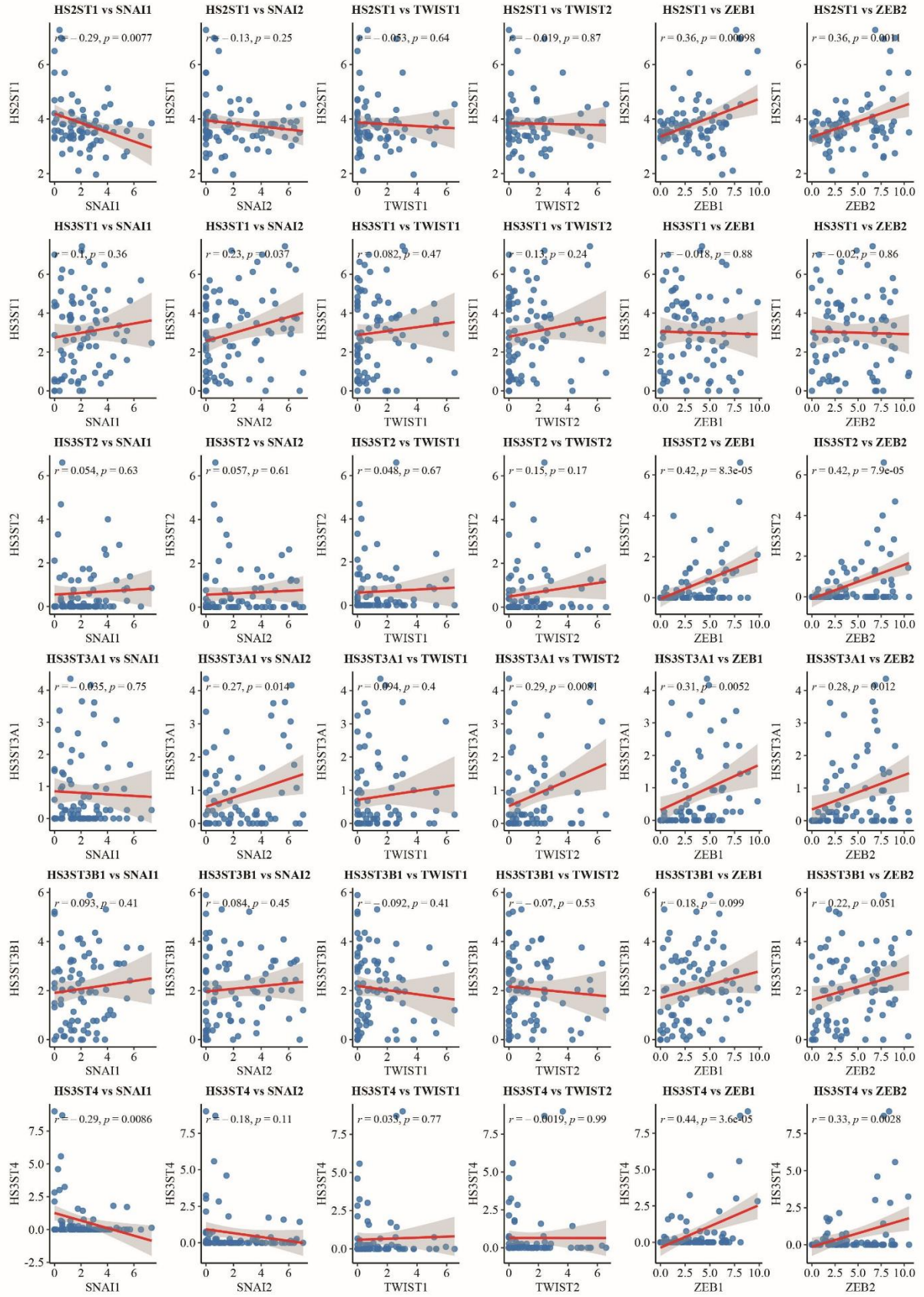


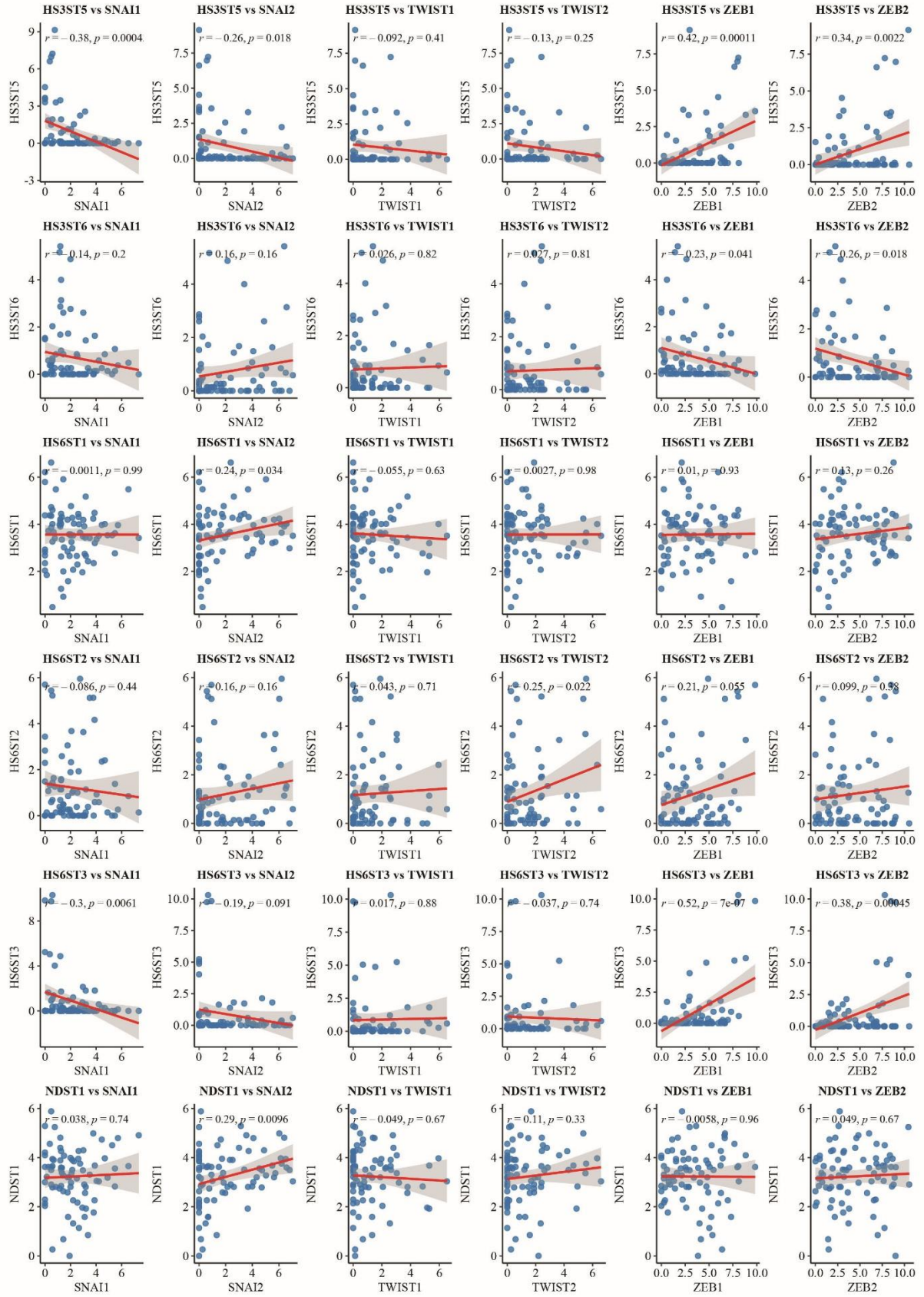


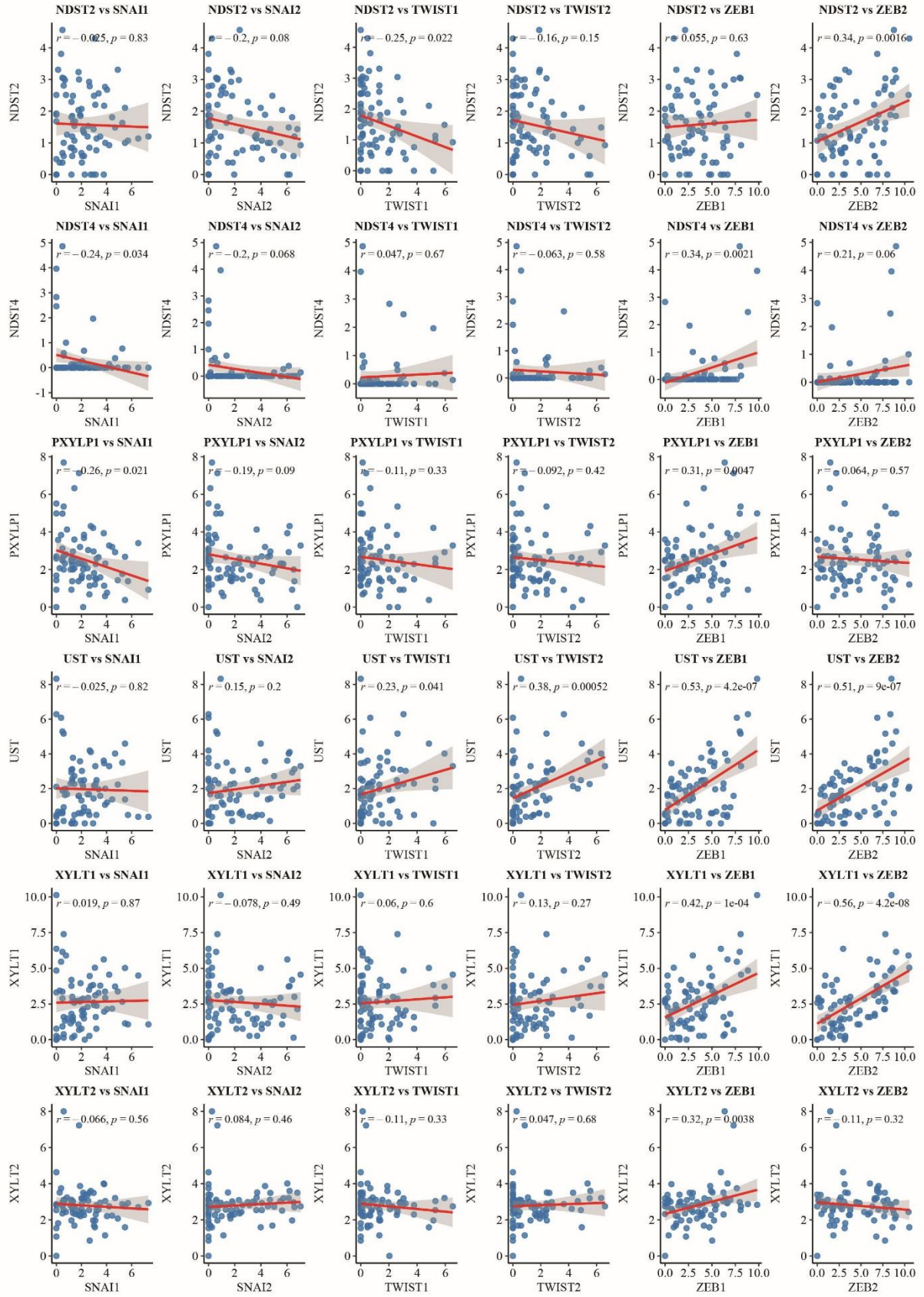




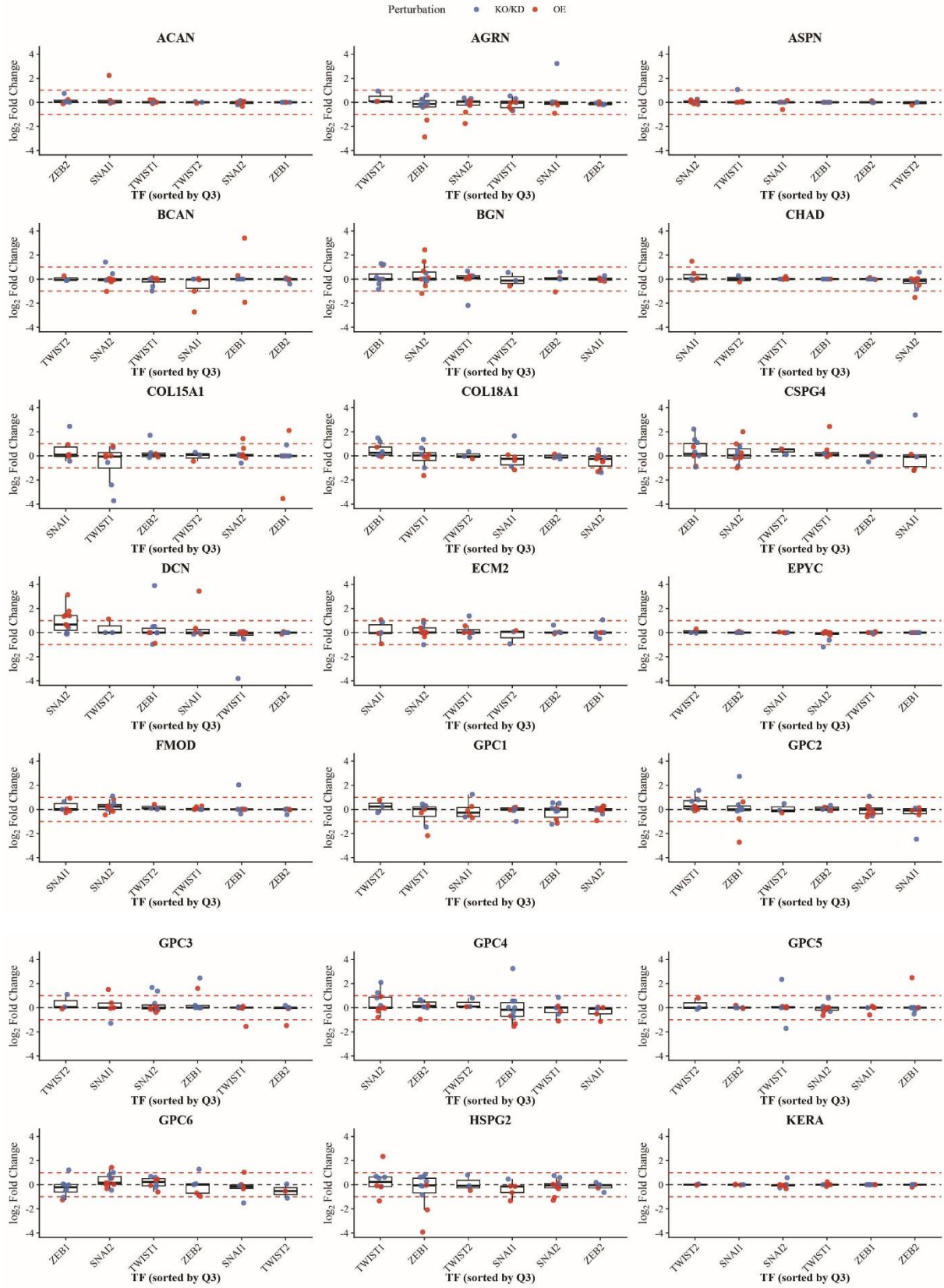


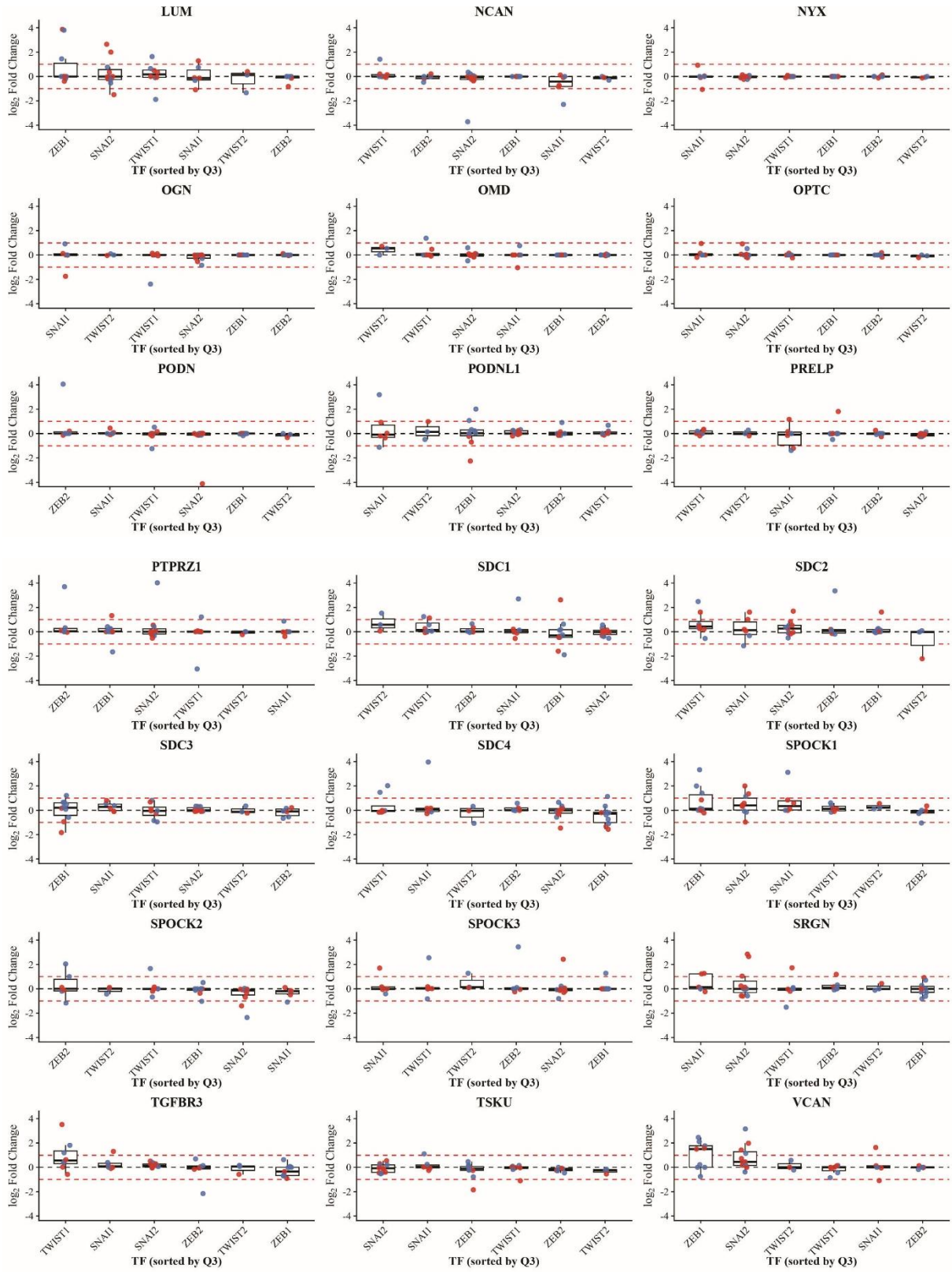




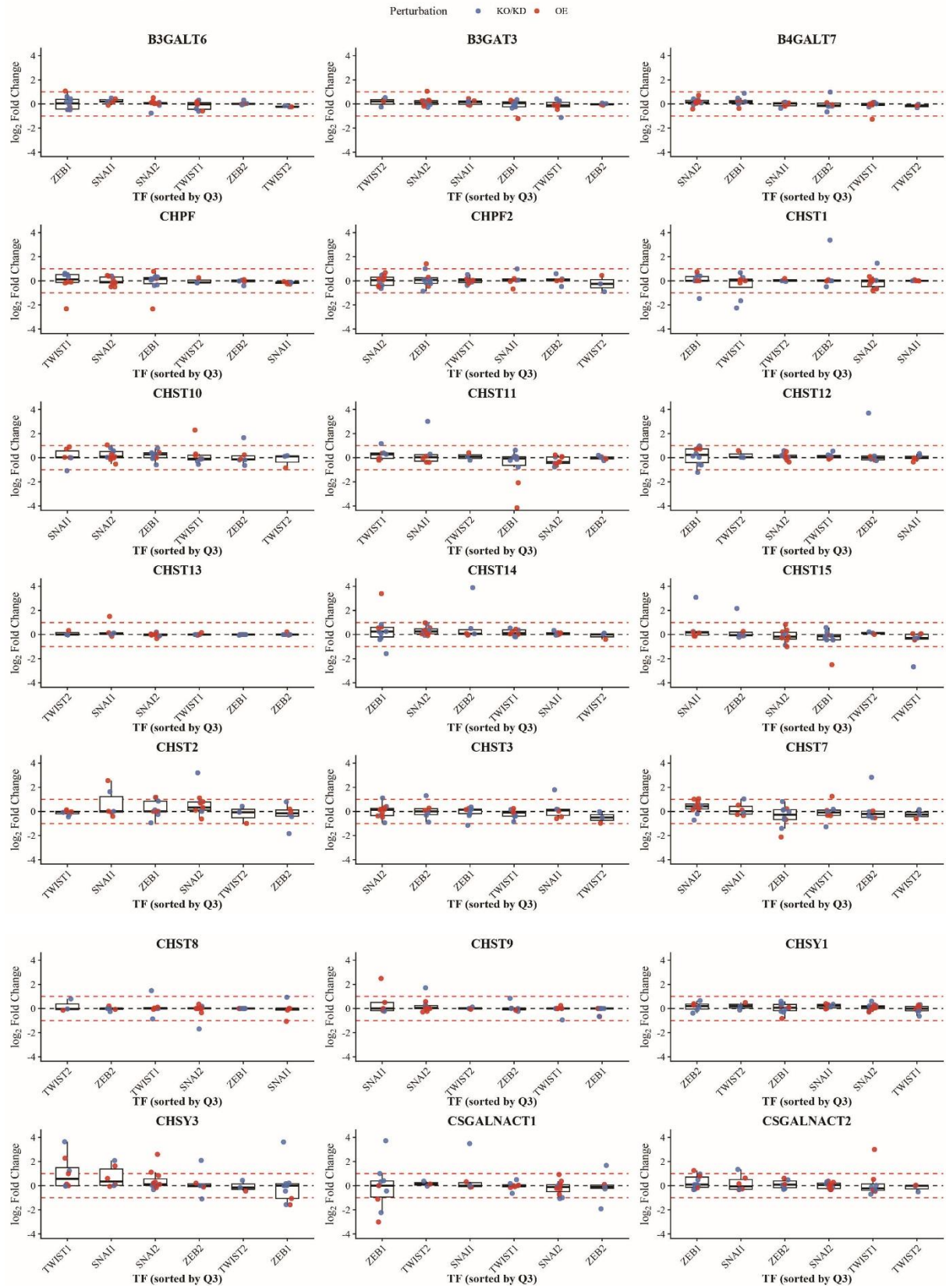


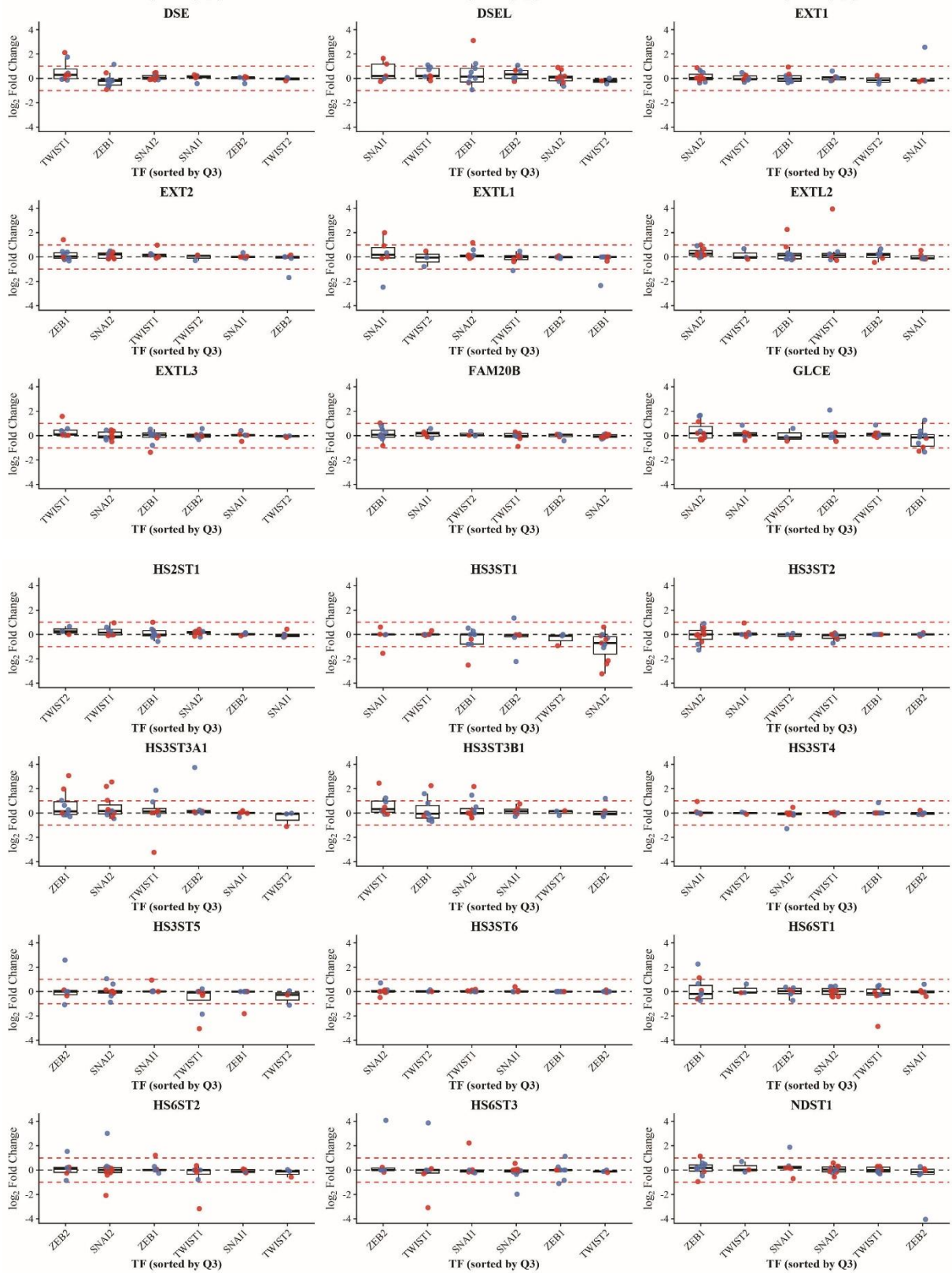
B Effect of EMT TF perturbation on proteoglycan core proteins - related genes

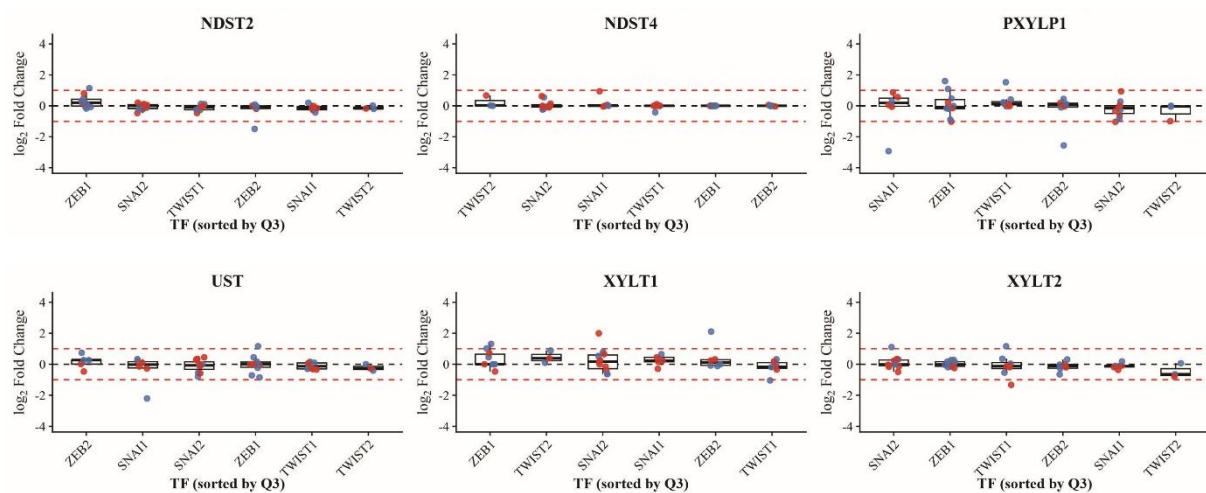




Effect of EMT TF perturbation on GAG biosynthetic enzymes-related genes







Supplementary Figure 3: Association and perturbation responses of proteoglycan and GAG genes to EMT transcription factors. (A) Pearson correlations between individual proteoglycan and GAG-pathway genes and EMT transcription factors across all cell types from adult tissue. (B) Boxplots showing response of EMT transcription factor perturbation experiments responses of proteoglycan and GAG genes. EMT TFs on the x-axis are ordered by their Q3(log₂FC) across perturbation datasets. For knockout/knockdown (KO/KD) experiments, log₂FC signs were flipped to match activation direction; overexpression (OE) experiments kept their original sign. Dot color indicates whether the experiment was KO/KD or OE.