

Integrated gene expression and alternative splicing analysis in human and mouse models of Rett Syndrome.

Silvia Gioiosa^{1,6*}, Silvia Gasparini^{2,6}, Carlo Presutti³, Arianna Rinaldi^{3,4}, Tiziana Castrignanò⁵ & Cecilia Mannironi^{2*}

¹CINECA, SuperComputing Applications and Innovation Department, Via dei Tizii 6, 00185, Rome, Italy. ²Institute of Molecular Biology and Pathology, National Research Council, Rome, Italy. ³Department of Biology and Biotechnology “C. Darwin”, Sapienza University of Rome, Rome, Italy. ⁴Center for Research in Neurobiology “D. Bovet”, ⁵Department of Ecological and Biological Sciences (DEB), University of Tuscia, Largo Università snc, 01100, Viterbo, Italy. ⁶ These authors contributed equally: Silvia Gioiosa and Silvia Gasparini.

¹CINECA, SuperComputing Applications and Innovation Department, Via dei Tizii 6, 00185, Rome, Italy. ²Institute of Molecular Biology and Pathology, National Research Council, Rome, Italy. ³Department of Biology and Biotechnology “C. Darwin”, Sapienza University of Rome, Rome, Italy. ⁴Center for Research in Neurobiology “D. Bovet”, ⁵Department of Ecological and Biological Sciences (DEB), University of Tuscia, Largo Università snc, 01100, Viterbo, Italy. ⁶ These authors contributed equally: Silvia Gioiosa and Silvia Gasparini.

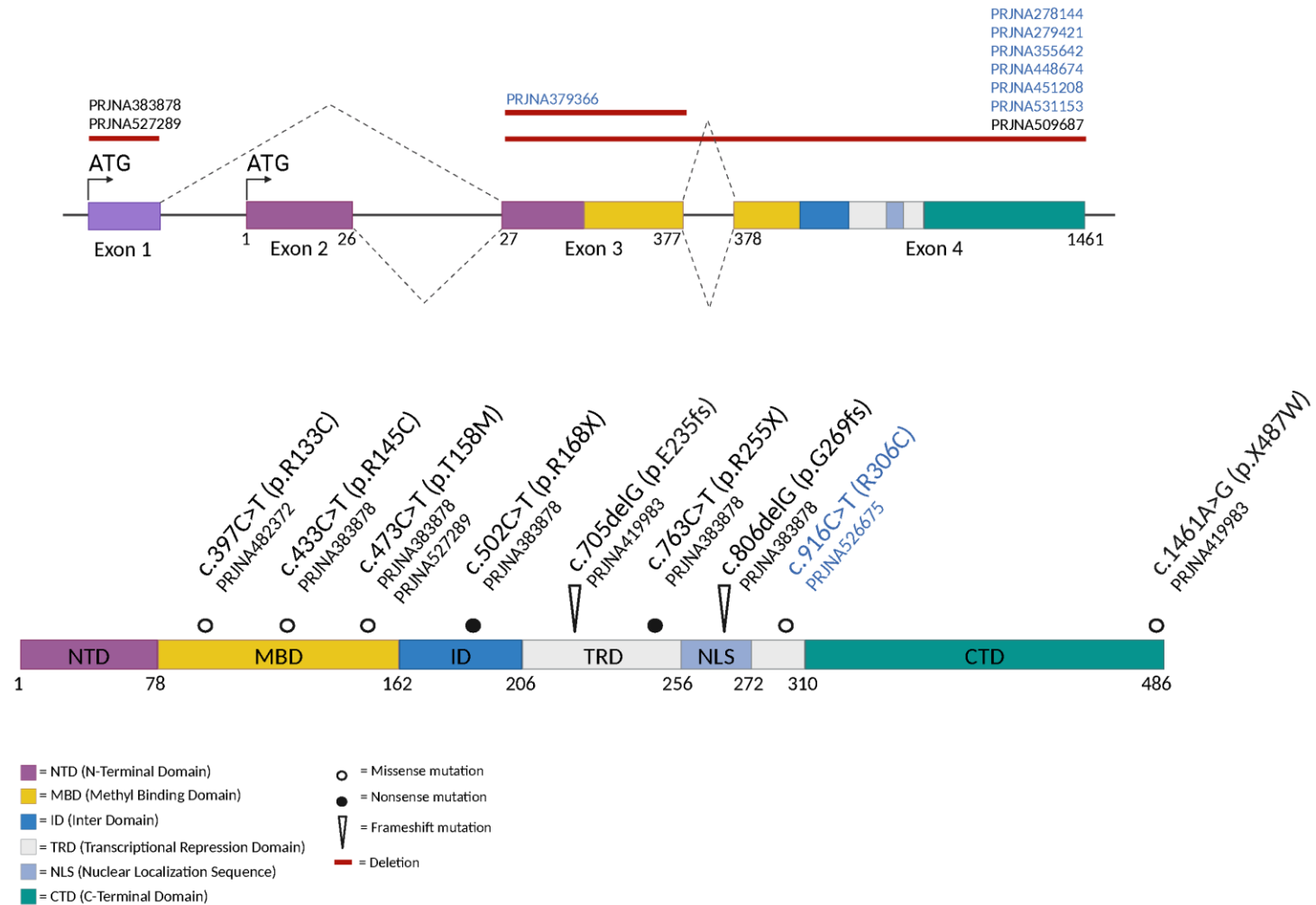


Figure S1. Graphical representation of MeCP2 gene and protein. Proteins domains and disease-causing mutations analyzed in this study are reported. Exon deletions are schematically represented at the gene level as red lines; frameshift, missense, and nonsense mutations are indicated above protein scheme. Human and mouse bioprojects are reported in black and blue, respectively. *Created with BioRender.com*

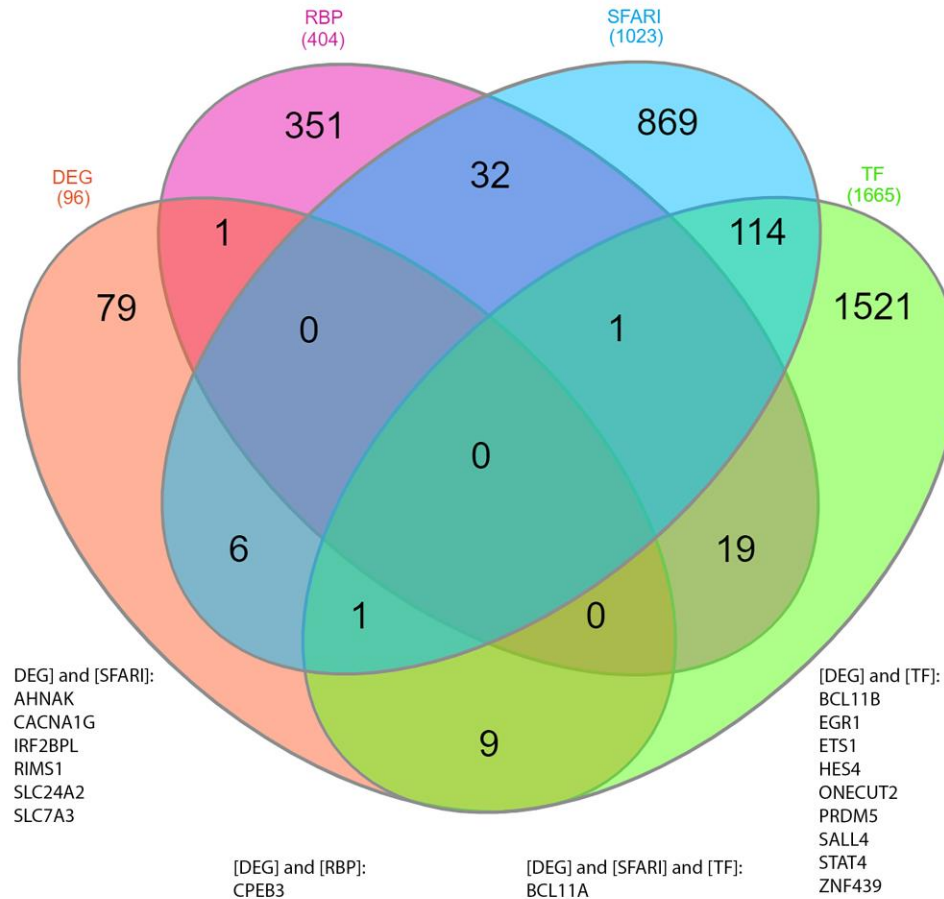


Figure S2. Sfari genes, RNA Binding Proteins , and Transcription factors among human common DEGs. The Venn diagram shows the overlap between RNA Binding Proteins (RBPs), Sfari genes, Transcription factors (TFs) and human RTT DEGs. The analysis was done on mouse DEGs shared by 3 or more datasets. The dimension of each list is reported on the top of the graph. See Additional file 2-Supplementaru Table 1 for Sfari, RBP and TF gene lists.

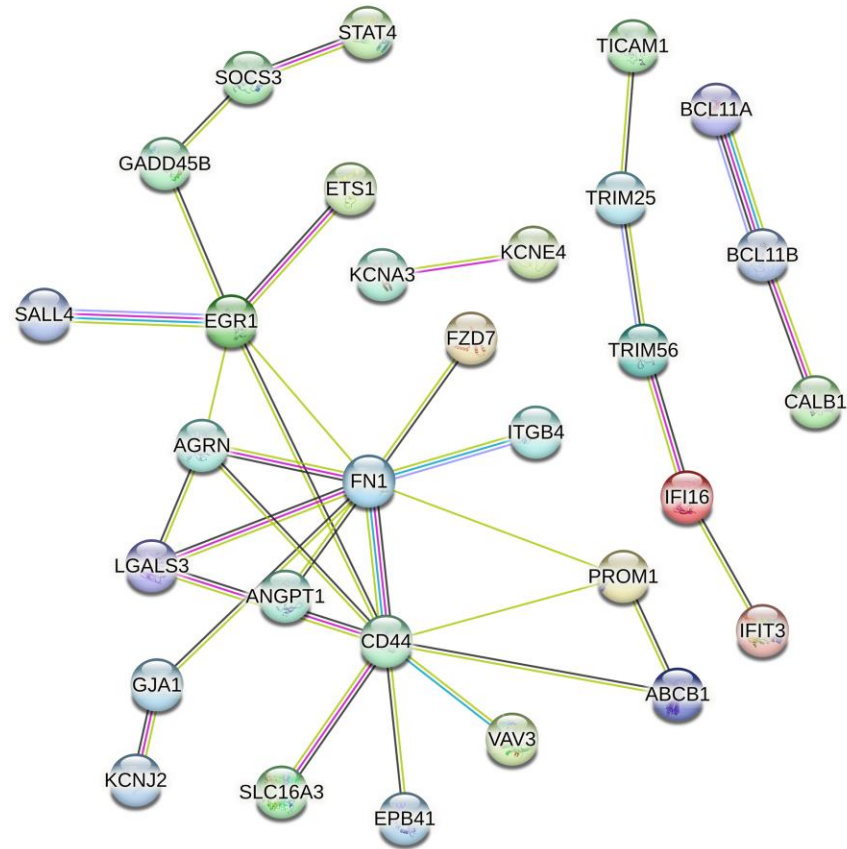


Figure S3. Protein-protein interaction (PPI) network of human common DEGs. PPI network was identified by String analysis on 96 common human DEGs. 88 nodes, 33 edges (expected number of edges:20) with a PPI enrichment p-value of 0.00514 have been identified. Average node degree: 0.75; Average local clustering coefficient:0.228; Meaning of network edges: evidence. Active interaction sources: textmining, experiments, databases, co-expression. Minimum required interaction score (confidence): 0.500. Nodes represent proteins, the network edges represent protein-protein association, known from curated databases (light blue), experimentally determined (violet), predicted from gene co-expression (black), or from text mining (yellow).

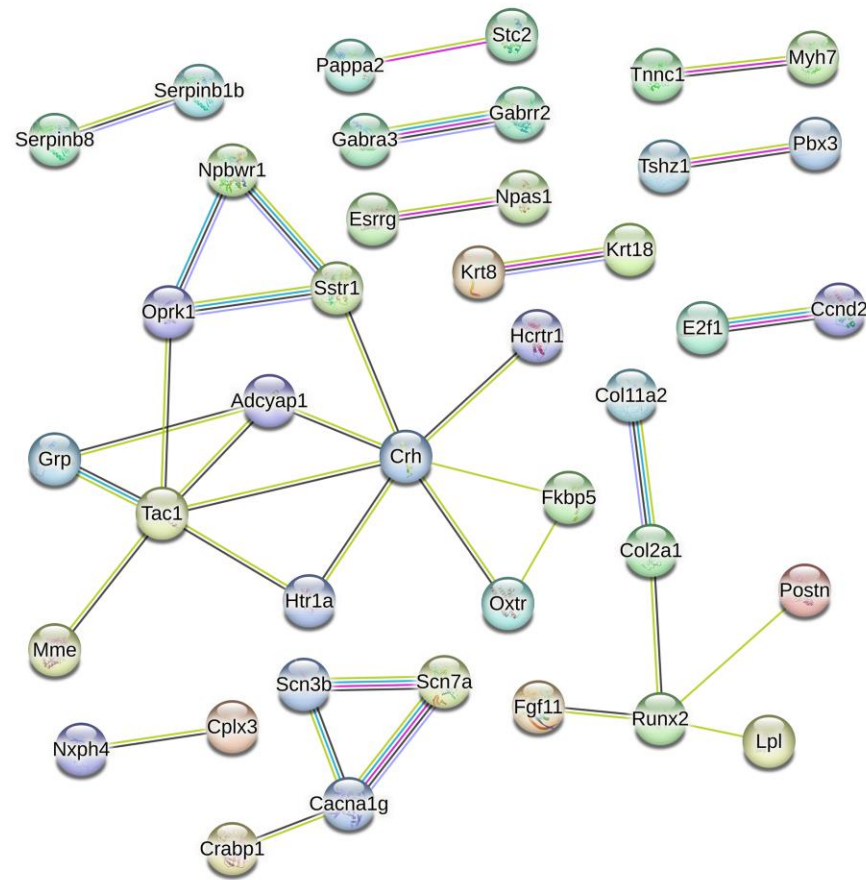


Fig. S5. Protein-protein network analysis of mouse common DEGs. PPI network was identified by String analysis on 113 mouse DEGs shared by at least 3 datasets. 109 nodes, 35 edges (expected number of edges: 11) with a PPI enrichment p-value of 5E-09 have been identified. Average node degree: 0.642; Average local clustering coefficient: 0.312; Meaning of network edges: evidence. Active interaction sources: textmining, experiments, databases, co-expression. Minimum required interaction score (confidence): 0.500. Nodes represent proteins, the network edges represent protein-protein association, known from curated databases (light blue), experimentally determined (violet), predicted from gene co-expression (black), or from text mining (yellow).

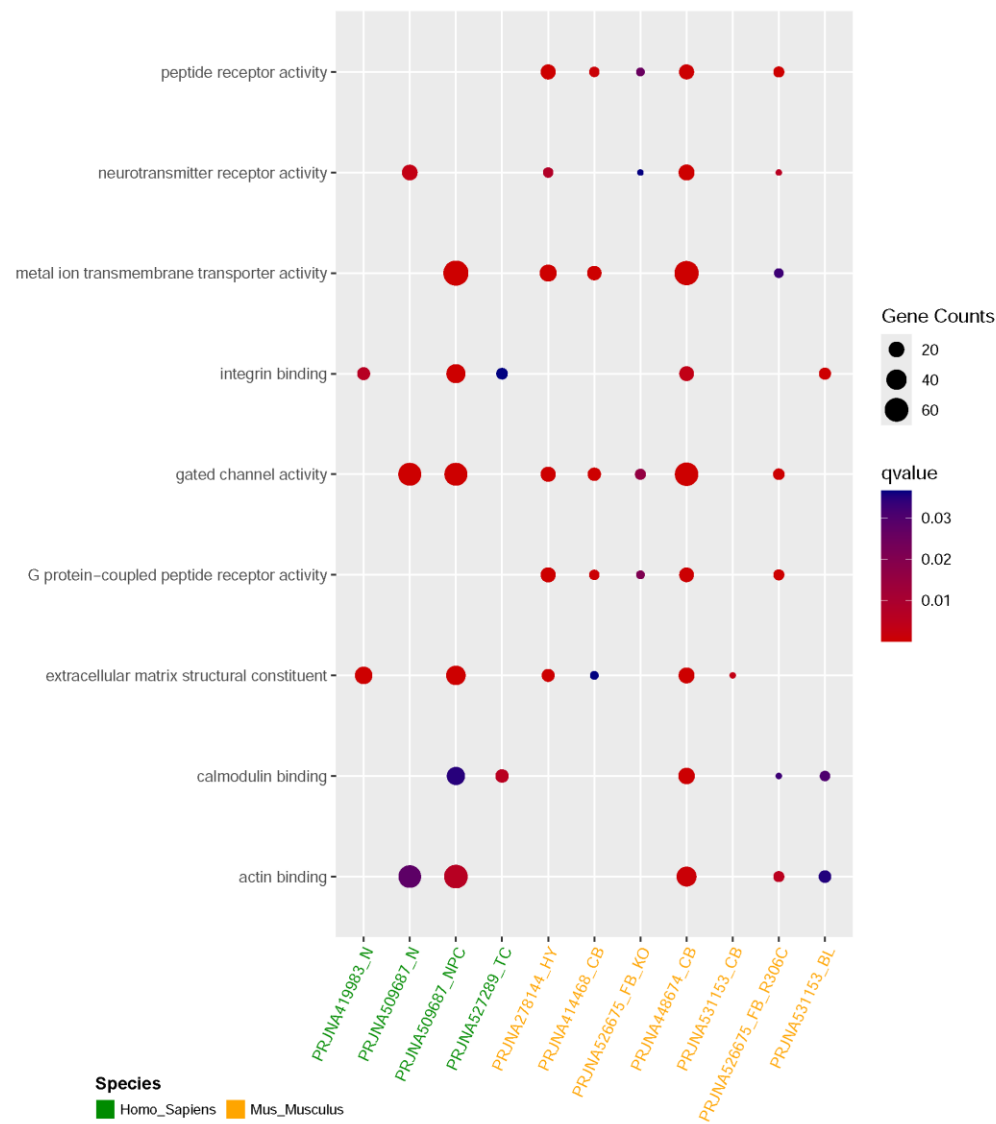
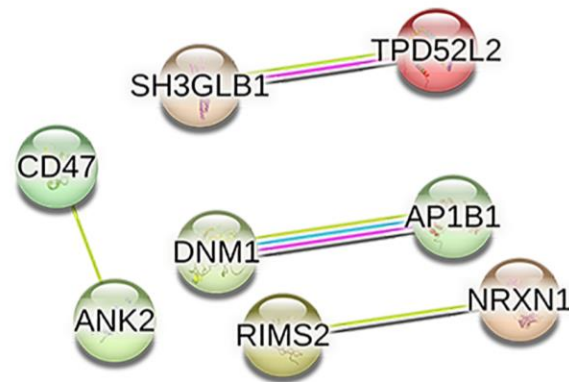


Figure S6 Human and Mouse DEG GO Functional Analysis. Compare Clusters-Simplify Analysis of Enriched GO Terms were identified by a Compare Cluster analysis on human and mouse common DEGs. The datasets enriched in common GO terms are reported on the lower side of the graph.



Term Description	Category	Observed Gene Count	Background Gene Count	Strength	q-value	Matching Proteins
Vesicle	GO Component	12	3957	0.52	0.0487	MAN2A1,DST,MINK1,ANK2,AP1B1,PTPRF,CD47,DNM1,NRXN1,RIMS2,SH3GLB1,HSPH1
Cell junction	UniProt Keywords	6	809	0.91	0.0398	ABI2,DST,MINK1,ANK2,NRXN1,RIMS2

Figure S7. Protein-protein network and functional analysis of human common DAS genes. String analysis was performed on human DAS shared by at least 3 datasets. 18 nodes, 4 edges (expected number of edges: 1), with a PPI enrichment p-value of 0.00411. Average node degree: 0.444, Average local clustering coefficient:0.444; Meaning of network edges: evidence; active interaction sources: textmining, experiments, databases, co-expression; minimum required interaction score (confidence): 0.500. Nodes represent proteins, the network edges represent protein-protein association, known from curated databases (light blue), experimentally determined (violet), predicted from gene co-expression (black), or from text mining (yellow) (**upper panel**); functional enrichment for the 'Vesicle' GO component and 'Cell junction' UniProt Keywords was identified in the PPI network of common human DAS genes (**lower panel**) .