The American Journal of Human Genetics, Volume *101*

Supplemental Data

DOMINO: Using Machine Learning to Predict Genes Associated with Dominant Disorders

Mathieu Quinodoz, Beryl Royer-Bertrand, Katarina Cisarova, Silvio Alessandro Di Gioia, Andrea Superti-Furga, and Carlo Rivolta

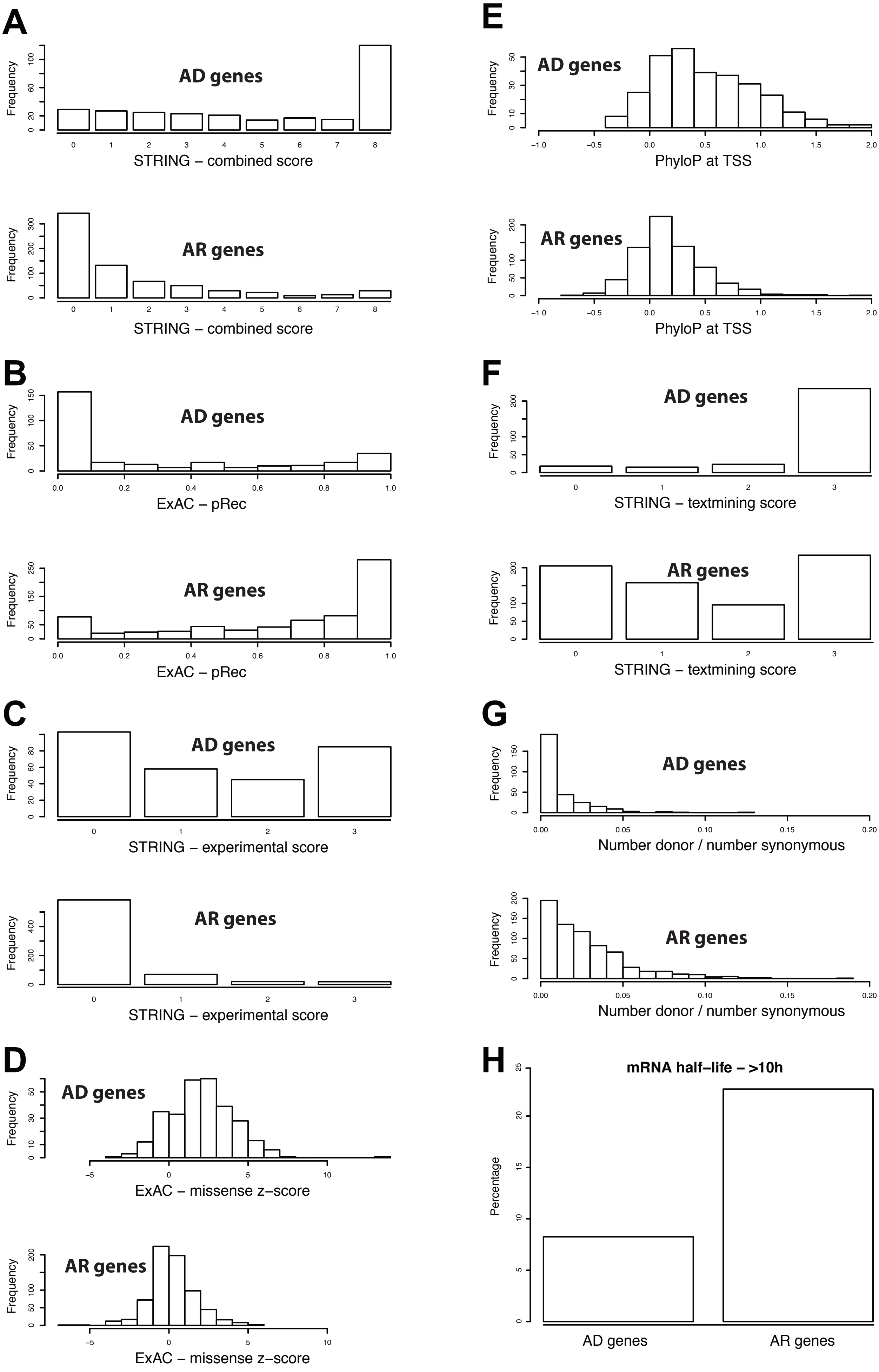

Figure S1. Histograms of selected features for AD and AR genes of the training set. (A) STRING – combined score. (B) ExAC – pRec. (C) STRING – experimental score. (D) ExAC missense z-score. (E) Average PhyloP at TSS. (F) STRING – text mining score. (G) Number of donor site variant / number of synonymous variant in ExAC. (H) Percentage of genes with an mRNA half-life >10h in mouse embryonic stem cells.

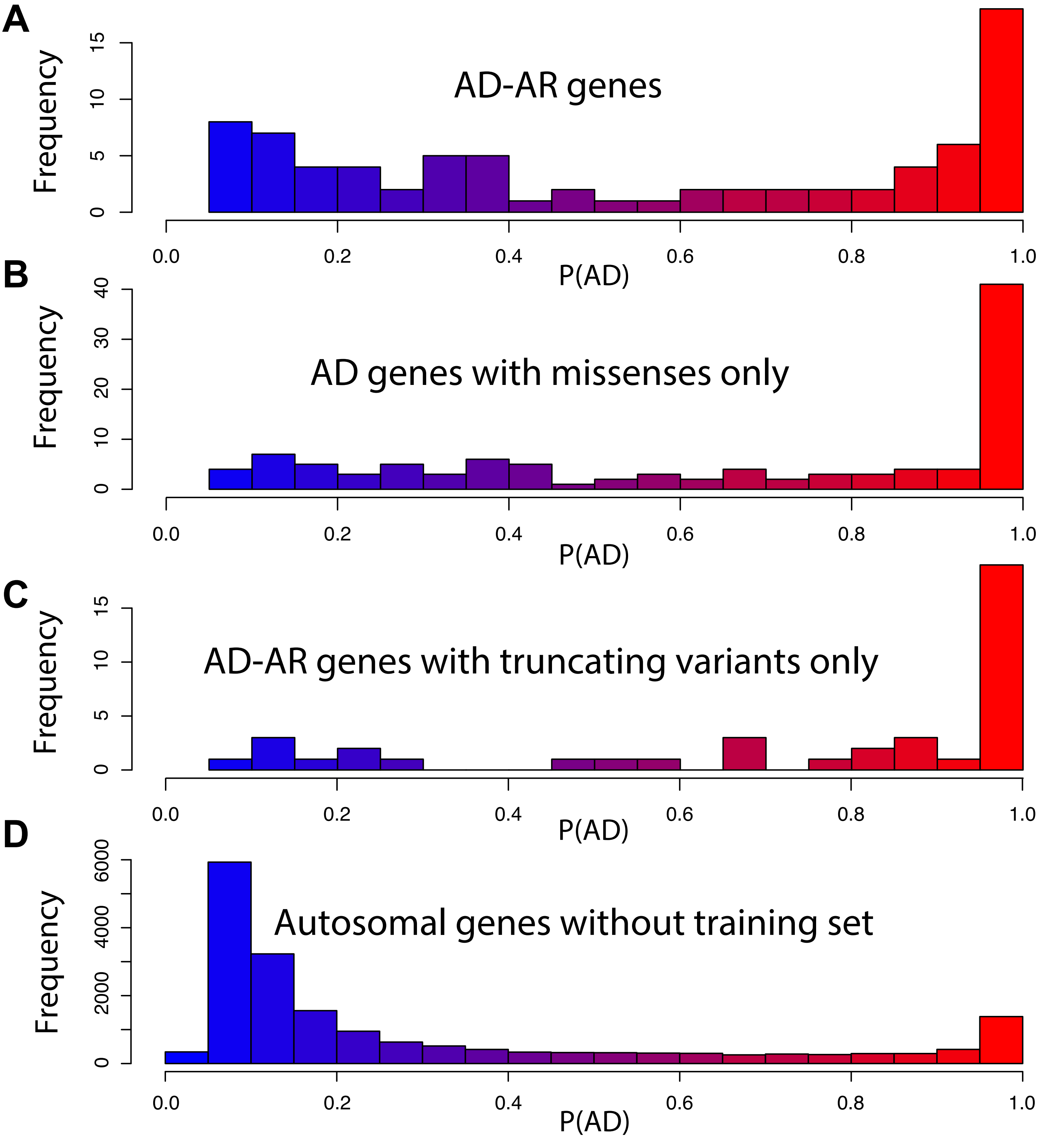

Figure S2. Histograms of P(AD) for specific gene categories. (A) Genes associated to both dominant and recessive inheritance of pathogenic traits. (B) Genes with only pathogenic missense. (C) Genes with only truncating pathogenic mutations. (D) Autosomal genes that were not in the training set (*N*=18,360).