Transposable Element Content Analysis

mRNA sequences for all relevant genes were retrieved from the UCSC Genome Browser using the Table Browser functionality. Promoter sequences were defined as all genomic content 2000 nucleotides upstream of the transcription start site. FASTA files containing either mRNA or promoter sequences were cleared of duplicates and renamed using a custom *Python* script. All resulting FASTA files were processed using *RepeatMasker* [] with the following option flags: *-no\_is -nolow -s -species Human -pa 8*. Parsing of *RepeatMasker* output was performed with bash command line tools and custom *Python* scripts. *SalmonTE* [] with default settings was used to quantify Transposable Element transcript expression from aggregate bulk RNA sequencing data and output was parsed using *R*. Analysis of all data was performed and visualized in *R* using custom scripts implementing the *Tidyverse* package.

Zinc Finger Protein Analysis

ChIP-exo data and supplementary information were extracted from supplementary data provided by *Imbeault et al* []. ZNF genes were cross referenced with *DESeq2* and *RepeatMasker* outputs to extract relevant differential expression data of ZNF proteins and Transposable Element transcripts using *R*. *RepeatMasker* output from promoter analyses was cross referenced with ChIP-exo target data to identify potential regulatory targets of differentially expressed KZNFs. Only KZNF targets with ‘score’ [see Imbeault *et al*] >= 75 were kept for analysis. Analysis of all data was performed and visualized in *R* using custom scripts implementing the *Tidyverse* package.

Gene Set Enrichment Analysis

*DESeq2* output was first ‘pre-ranked’ by *log2FoldChange* in *R* and then loaded into the *javaGSEA* desktop application. The *GSEAPreranked* functionality with default settings was used along with all *MSigDB* Hallmark gene sets to determine enrichment. *.xls* output was parsed and visualized in *R* using scripts implementing the *Tidyverse* package.

Gene Ontology Analysis

Upregulated gene names were extracted from *DESeq2* output using bash command line tools. Name lists were pasted into the *Gene Ontology Consortium*’s *Enrichment Analysis* tool powered by *PANTHER*. Output data was exported as *.txt* files and parsed using bash command line tools. Parsed data was visualized using custom *R* scripts.

Smit, AFA, Hubley, R & Green, P. *RepeatMasker Open-4.0*.  
2013-2015 <http://www.repeatmasker.org>.

SalmonTE -- Guo, C., Jeong, H.-H., Hsieh, Y.-C., Klein, H.-U., Bennett, D.A., Jager, P.L.D., Liu, Z., and Shulman, J.M. (2018). Tau Activates Transposable Elements in Alzheimer’s Disease. Cell Reports *23*, 2874–2880.

Imbeault, M., Helleboid, P.-Y., and Trono, D. (2017). KRAB zinc-finger proteins contribute to the evolution of gene regulatory networks. Nature *543*, 550.

R Core Team (2018). R: A language and environment for statistical computing. R

Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

Hadley Wickham (2017). tidyverse: Easily Install and Load the 'Tidyverse'. R package

version 1.2.1. <https://CRAN.R-project.org/package=tidyverse>

GSEA -- Subramanian, Tamayo, et al. 2005 Proc Natl Acad Sci U S A 102(43):15545-50

MSigDB -- Liberzon et al. 2011 Bioinformatics 27(12):1739-40

ALL GO REFS:

Ashburner et al. Gene ontology: tool for the unification of biology (2000) *Nat Genet* **25(1)**:25-9. [Online at Nature Genetics](https://www.ncbi.nlm.nih.gov/pubmed/10802651).

GO Consortium, Nucleic Acids Res., 2017

Mi et al., Nucleic Acids Res., 2017