

## Supplementary Information

### **Genes essential for embryonic stem cells are associated with neurodevelopmental disorders**

Shahar Shohat and Sagiv Shifman

#### **Supplementary data:**

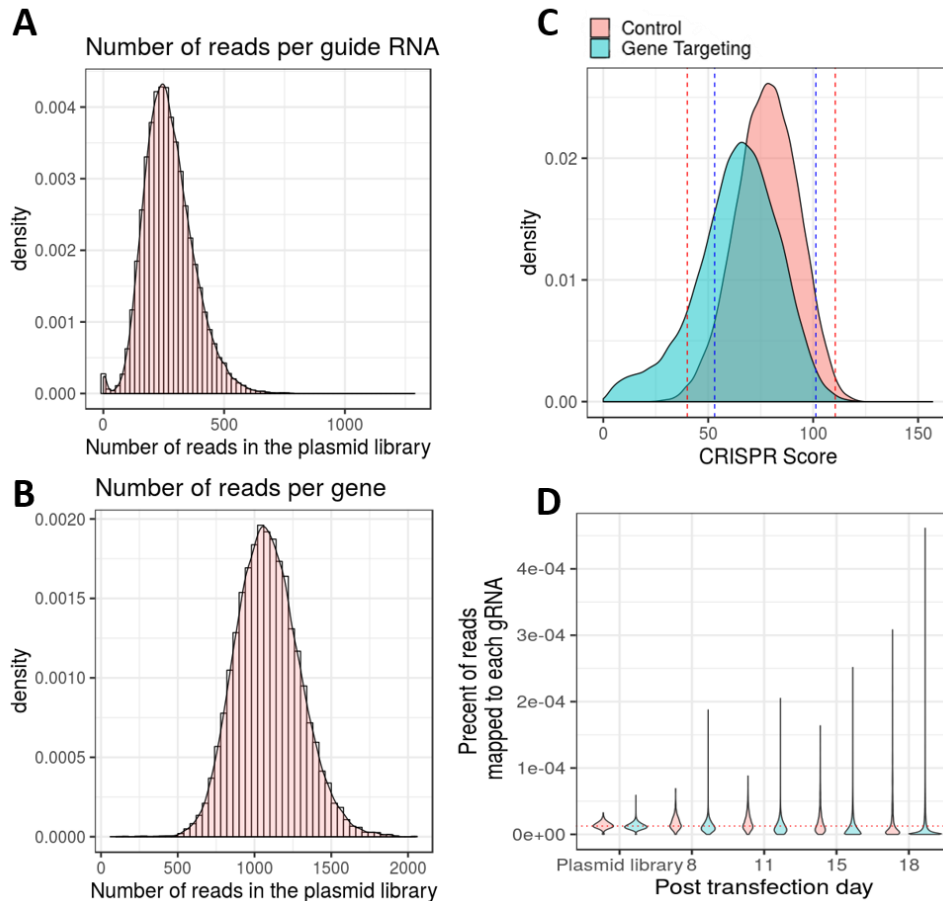
Figure S1 (Page 1)

Figure S2 (Page 2)

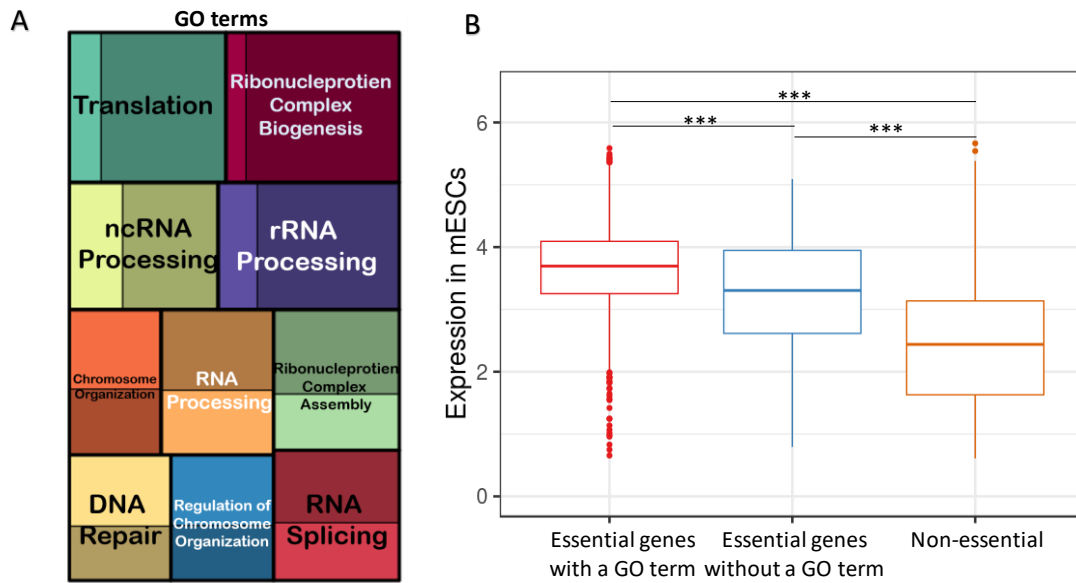
Figure S3 (Page 3)

Figure S4 (Page 4)

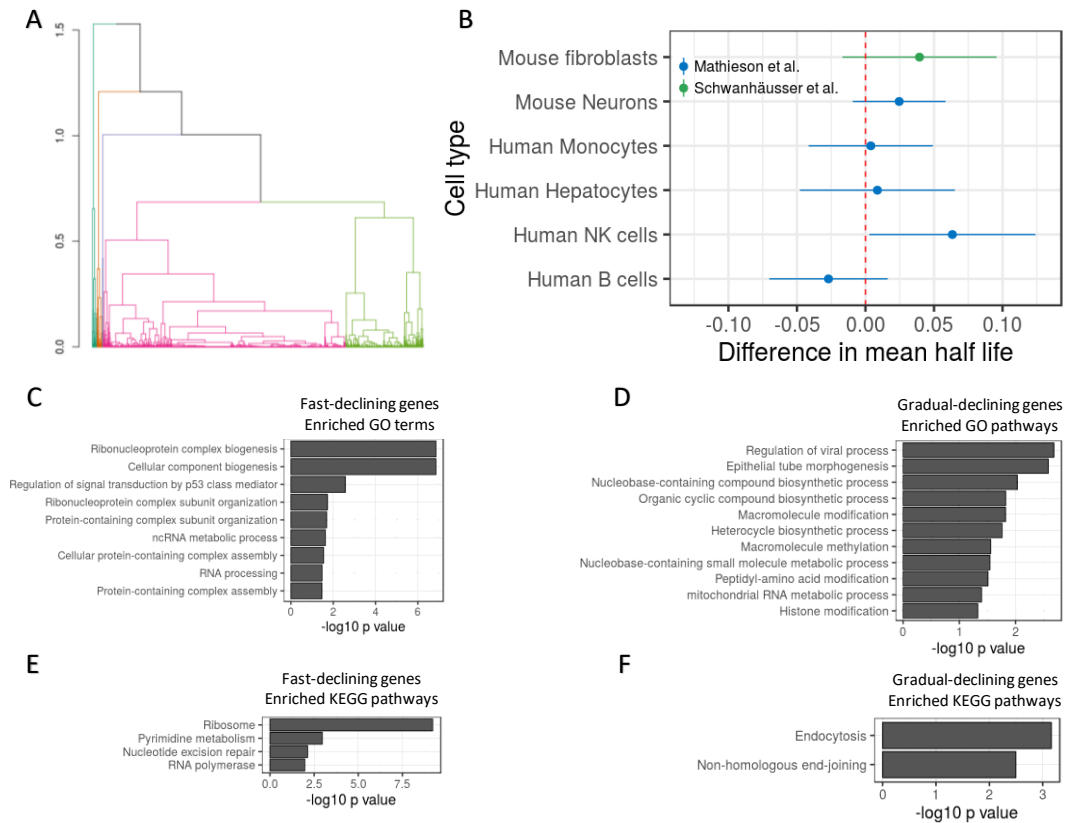
Figure S5 (Page 5)



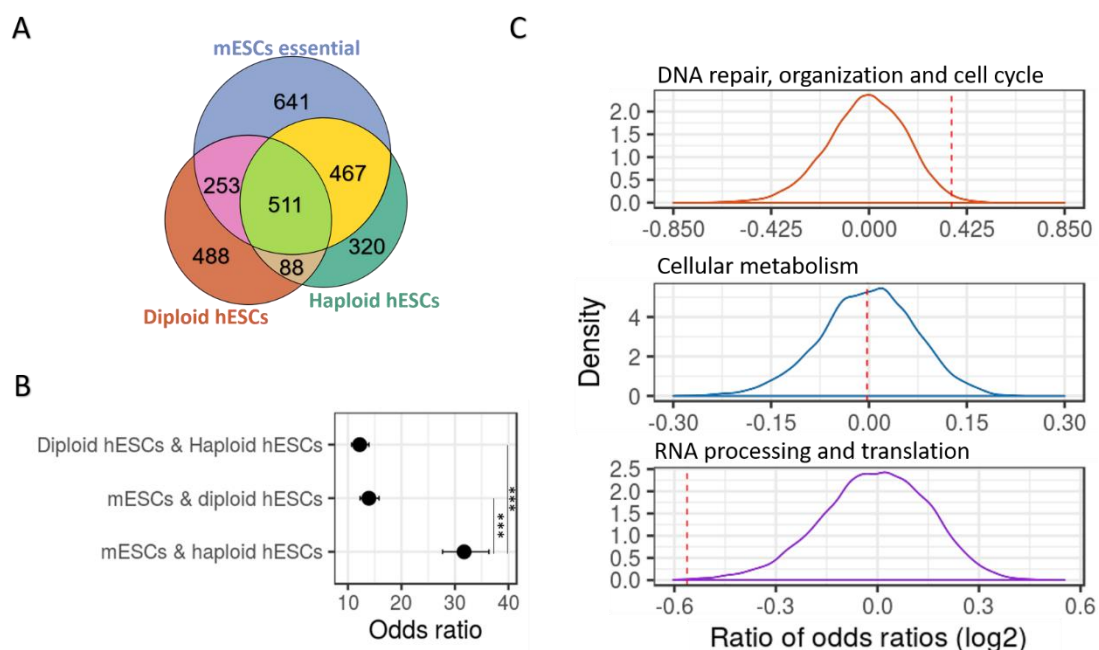
**Figure S1. Analysis and performance of the CRISPR pooled library.** **(A)** Distributions of the number of reads per gRNA **(B)** Distributions of the number of reads per gene. **(C)** The distribution of CRISPR scores for gRNAs targeting genes (blue) and simulated genes based on control gRNAs (pink). The dashed vertical lines represent nominal (blue) and FDR corrected (red)  $P = 0.05$ . **(D)** Violin plot showing the proportion of reads mapped to each gRNA for gRNAs targeting genes (blue) and control gRNAs (pink) across different days *in vitro*.



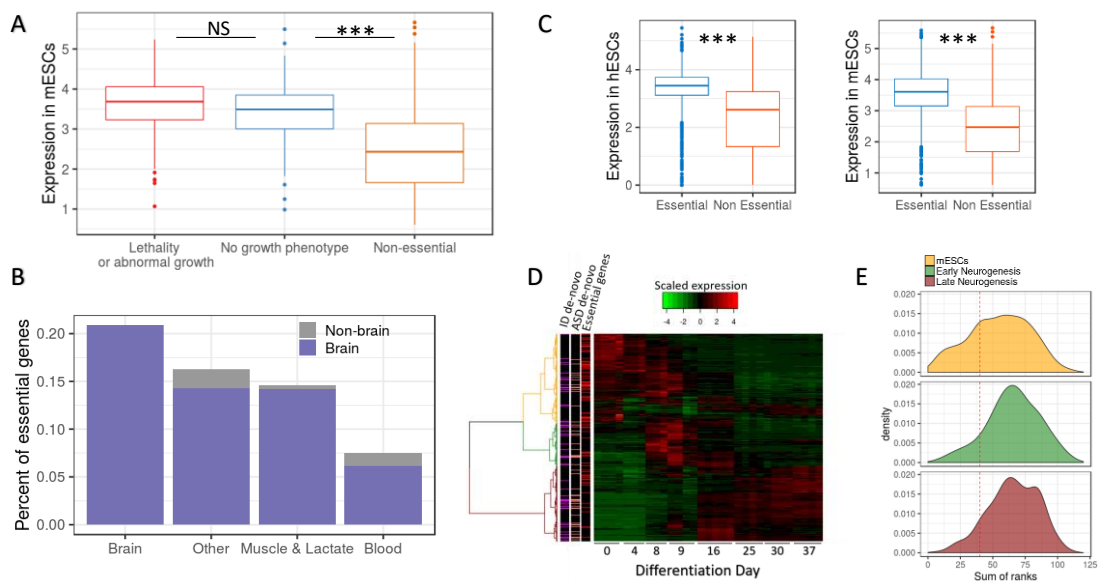
**Figure S2. Gene ontologies associated with essential genes. (A)** A treemap of the top 10 most enriched GO terms. The square size is proportional to the enrichment strength and the color intensity indicates the proportion of essential (dark color) and non-essential genes (bright colors) in the term. **(B)** Expression in mESCs of essential genes associated with a GO term, not associated with any GO term, and non-essential genes. Expression values are signal intensities ( $\log_{10}$ ) from microarray data. \*\*\*,  $P < 0.001$ .



**Figure S3. Dynamics of gRNA depletion.** (A) Hierarchical clustering of essential genes according to the rate of gRNA depletion. The red cluster contains the fast-declining genes and the light green cluster the gradual declining genes. (B) Non-significant differences in mean protein half-life between genes in the fast and gradual declining groups. Values are the mean difference in half-life ( $\log_{10}$  hours) between genes in the fast and gradual declining groups. Error bars indicates 95% confidence interval. (C-F) GO and KEGG enrichment analysis for genes in the fast and gradual declining groups.



**Figure S4. Overlap between genes essential in mESCs and genes essential in hESCs.** (A) Venn diagram displaying the overlap between essential genes in mESCs, haploid hESCs and diploid hESCs. (B) Pair-wise analysis of the overlap of essential genes between different ESCs. Values are odds ratio  $\pm$ 95% confidence interval. (C) (Related to 4B) Distribution of the ratios of odds ratios in simulated data under the null hypothesis of no association between ESCs essential genes and the groups of related GO terms. The red line shows the observed ratio of odds ratios in the real data. \*\*\*\*,  $P < 0.001$ .



**Figure S5. Phenotypes associated with essential genes. (A)** The box plots show the expression levels in mESCs of three groups of genes: (1) essential genes associated with lethality or abnormal growth in mice, (2) essential genes not associated with growth phenotypes, and (3) non-essential genes. Values are signal intensities ( $\log_{10}$ ) from a microarray data. **(B)** Percentage of essential genes associated with brain, muscle & lactate, blood or other phenotypes. Color indicates the fraction of genes included also in brain related phenotypes. **(C)** Gene expression in hESCs (left) and mESCs (right) for genes essential in ESCs relative to non-essential genes. For hESCs the values are normalized counts from RNA-Seq ( $\log_{10}$ ) for mESCs values are signal intensities ( $\log_{10}$ ) from a microarray data. **(D)** A heatmap of gene expression of risk genes for neurodevelopmental disorders during mouse *in vitro* corticogenesis. The color of the heatmap correspond to the normalized expression (red is high and green is low levels of expression). Side bars indicate ESCs essential genes (red), genes implicated in ASD by multiple *de novo* mutations (salmon) and genes implicated in ID by multiple *de novo* mutations (magenta). **(E)** Distribution of CRISPR scores for NDD risk genes divided by expression patterns during mouse *in vitro* corticogenesis. The red line indicates a corrected  $P < 0.05$ . NS, non-significant; \*\*\*  $P < 0.001$ .