Co-expression enrichment analysis at the single-cell level reveals convergent defects in neural progenitor cells and their cell-type transitions in neurodevelopmental disorders

Kaifang Pang, Li Wang, Wei Wang, Jian Zhou, Chao Cheng, Kihoon Han, Huda Y. Zoghbi & Zhandong Liu

Supplemental Methods

Controlling for confounding factors in co-expression enrichment analysis

In addition to using the threshold of top 0.5% to construct co-expression networks and calculate co-expression fold enrichment score for NDD gene sets in six major cell types, we used different thresholds of top 0.25% and top 1%. We also varied the thresholds between top 0.1% and top 5% to construct co-expression networks and calculate co-expression fold enrichment score for dnLoF-ASD genes in NPCs and dnMis-Epi genes in interneurons. In addition to using Spearman's correlation to construct co-expression networks and calculate co-expression fold enrichment score at the threshold of top 0.5% for dnLoF-ASD genes in NPCs and dnMis-Epi genes in interneurons, we used another 16 measures of association implemented in the 'dismay' R package (Skinnider et al. 2019). Moreover, we assessed the effect of gene set size difference on the co-expression fold enrichment score of NDD and control gene sets in six major cell types. For each major cell type, we first determined the smallest gene set size of NDD and control gene sets with genes expressed in that type. We then downsampled the same number of genes (the smallest gene set size) 1000 times for each gene set to calculate the co-expression fold enrichment score. We also evaluated the dependence of gene expression on the co-expression fold enrichment score of NDD gene sets in six major cell types. For each major cell type, genes were divided into ten bins based on expression level, with each bin containing equal numbers of genes. For each gene set in each cell type, the co-expression enrichment score was computed using 1000 randomly chosen

same-size gene sets with the same expression distribution across bins in the cell type as the background gene set.

CHD8 target gene analysis

The analytic results of *Chd8* haploinsufficient mice RNA-seq data were obtained from Supplemental Table S3 in (Gompers et al. 2017). Only genes in the *Chd8* RNA-seq data (with gene *CHD8* removed) that are also expressed in NPCs in the human cortical scRNA-seq data were defined as background genes for *CHD8* target gene analysis. We defined *CHD8*-activated and -repressed genes as the top 300 downregulated and top 300 upregulated genes, respectively, based on $\log_2$(fold change) values in *Chd8* haploinsufficient versus wild-type mice at each developmental stage. *CHD8*-bound genes are genes whose promoters are bound by *Chd8* in adult mouse forebrain using ChIP-seq (Gompers et al. 2017). To compute the overlap between *CHD8*-activated/-repressed genes and *CHD8*-bound genes (or ASD genes with at least one dnLoF mutation), we used only *CHD8*-bound (or ASD) genes that are also in the background gene set. P values of the overlap between *CHD8*-activated/-repressed genes and *CHD8*-bound (or ASD) genes were computed using the one-sided Fisher's exact test.

Based on the $\log_2$(fold change) value for any gene during the transition from vRG cells to IPCs at GW10, we obtained the distribution of $\log_2$(fold change) values for *CHD8*-activated or -repressed genes. Then, we computed P values that indicate whether *CHD8*-activated (-repressed) genes have higher (or lower) $\log_2$(fold change) values than the background genes during the transition by the one-sided Wilcoxon rank-sum test. Next, we computed the Spearman's correlation coefficient between any background gene and *CHD8* during the transition from vRG cells to IPCs at GW10 and obtained the distribution of correlations with *CHD8* for *CHD8*-activated or -repressed genes. We then computed P values which indicate whether *CHD8*-activated (-repressed) genes have higher (lower) correlations with *CHD8* than the background genes during the transition by the one-sided Wilcoxon rank-sum test. To compute the overlap between *CHD8*-activated/-repressed genes and GO biological process terms, we used only GO terms with the remaining gene number between 10 and 1000 after filtering by the

background genes. P values of the overlap between *CHD8*-activated/-repressed genes and GO terms were computed using the one-sided Fisher's exact test.

References

Gompers AL, Su-Feher L, Ellegood J, Copping NA, Riyadh MA, Stradleigh TW, Pride MC, Schaffler MD, Wade AA, Catta-Preta R, et al. 2017. Germline Chd8 haploinsufficiency alters brain development in mouse. *Nat Neurosci* **20**: 1062–1073.

Skinnider MA, Squair JW, Foster LJ. 2019. Evaluating measures of association for single-cell transcriptomics. *Nat Methods* **16**: 381–386. http://www.nature.com/articles/s41592-019-0372-4.