circRNA-producing gene function enrichment analysis

Functional enrichment analysis of circRNA host genes was performed using the C5 gene sets (GO terms) implemented in the MSigDB database (version v7.2) using Fisher's exact test. Each gene set contains genes annotated to the same GO term. For each gene set, the Fisher's exact test was performed for k, K, N - K, n; where k is the number of circRNA host genes that are part of a GO term gene set; K is the total number of genes annotated to the same GO term gene set; N is the total number of all annotated human genes in GENCODE (v19); and n is the number of genes in the query set. The top 10 GO terms in each GO category (BP, MF, and CC) enriched in these circRNA host genes are further slimmed using a similar algorithm as the clusterProfile package[1].

We also evaluated whether there is a specific enrichment among circRNAs in genes associated with brain disorders. We used diseases in MeSH C10 (Nervous System Diseases) or F03 (Mental Disorders) for brain disorders, and associated diseases to genes using GenDisNet database. The disease-gene association was extracted from DisGeNet[2] (http://www.disgenet.org/) filtered with GDA score > 0.1. For all annotated protein-coding genes, we performed Fisher's exact test based on whether a gene is associated with a brain disorder and a gene hosts a circRNA.

1. Yu, G., Wang, L.-G., Han, Y. & He, Q.-Y. clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *OMICS: A Journal of Integrative Biology* **16**, 284–287 (2012).

2. Piñero, J. *et al.* DisGeNET: A comprehensive platform integrating information on human disease-associated genes and variants. *Nucleic Acids Research* **45**, D833–D839 (2017).