Help

Why are expression levels for some genes unreliable?

1. Several transgenic worm strains were used to isolate neurons for scRNA-seq. The transgenes used to generate these strains often lead to overexpression of genes. The following genes show both ectopic expression and overexpression (OE): *unc-54* (the *unc-54* 3’ UTR is used for all but one of the strains used), *dpy-20, rol-6, lin-15A, lin-15B, unc-119, pha-1* (all full-length OE in rescue constructs), *F38B6.2, C30F8.3, cex-1, C30A5.16*, and *saeg-2* (regions of these genes are included in promoter sequences for some strains).

Five additional genes are over-expressed under the control of endogenous regulatory elements: *eat-4, cho-1, unc-53, unc-47* and *gcy-35.* These genes have inflated expression levels, but do not show ectopic expression.

There are also additional genes that are likely induced by dissociation-related cellular stress.

How are the displayed TPM expression values calculated?

1. The expression values are in transcripts per million, TPM, as calculated in Packer, et al., 2019 (<https://science.sciencemag.org/content/365/6459/eaax1971>). Please note this is a different version of TPM than commonly used in bulk RNA sequence analysis. This single-cell version has no gene length normalization. Raw UMI counts are first normalized by dividing by a cell-specific size factor. Normalized counts are then averaged across all the cells corresponding to each annotated cell type. The average value is divided by the sum of averaged expression values for each cell type, and this is multiplied by 1,000,000 to give the TPM value.

The proportion of cells expressing a gene is calculated as the percentage of individual cells corresponding to a given cell type with at least 1 UMI for the gene.

What is the thresholding procedure used?

1. For details on the thresholding procedure, please see the Methods in our preprint: <https://www.biorxiv.org/content/10.1101/2020.12.15.422897v2>

What are the differences between threshold levels?

1. Thresholding was used to determine if a gene was expressed in a cell or not. Genes below the threshold in a given neuron were set to 0 TPM. Genes above the threshold have expression values calculated as described above. We offer expression levels at four different thresholds. Threshold 1 is the least stringent and has the highest numbers of expressed genes per neuron type, at the risk of some possible false positives. Threshold 4 is the most stringent, with the fewest genes detected per neuron type. This reduces false positives at the cost of increased false negatives. We feel threshold 2 offers a good balance. See preprint (Methods and Figure S6) for more details. The true positive rates and false discovery rates for the 4 thresholds are:

|  |  |  |
| --- | --- | --- |
| Threshold | True Positive Rate | False Discovery Rate |
| 1 | 0.853 | 0.197 |
| 2 | 0.810 | 0.140 |
| 3 | 0.696 | 0.104 |
| 4 | 0.582 | 0.084 |