# TDP43 Cryptic Expression in LCM Neurons

Introduction

We wanted to know if we could detect TDP43 cryptic events in LCM neurons from ALS or control patients. AL sent me the counts of spliced reads supporting the inclusion of various TDP-43 related cryptic events from a 2018 study.

There are 21 unique samples and 770 splice junctions in the data.

The dataframe was extended to include counts for all samples and all junctions, inserting spliced reads counts as '0'. A new column was added to include information on the disease group for each sample (i.e., ALS patients or control samples).

```
group_column <- spliced_reads |>
  select(junction_name, sample_name, n_spliced_reads) |>
  complete(junction_name, sample_name,fill=list(n_spliced_reads = 0)) |>
  mutate(disease = ifelse(grepl("ALS", sample_name),
                          "ALS",
                          "control"))
```

To visualise the spliced reads data, a density plot was made with a pseudo log scale on the x-axis to compare the number of spliced reads in the ALS and control samples.
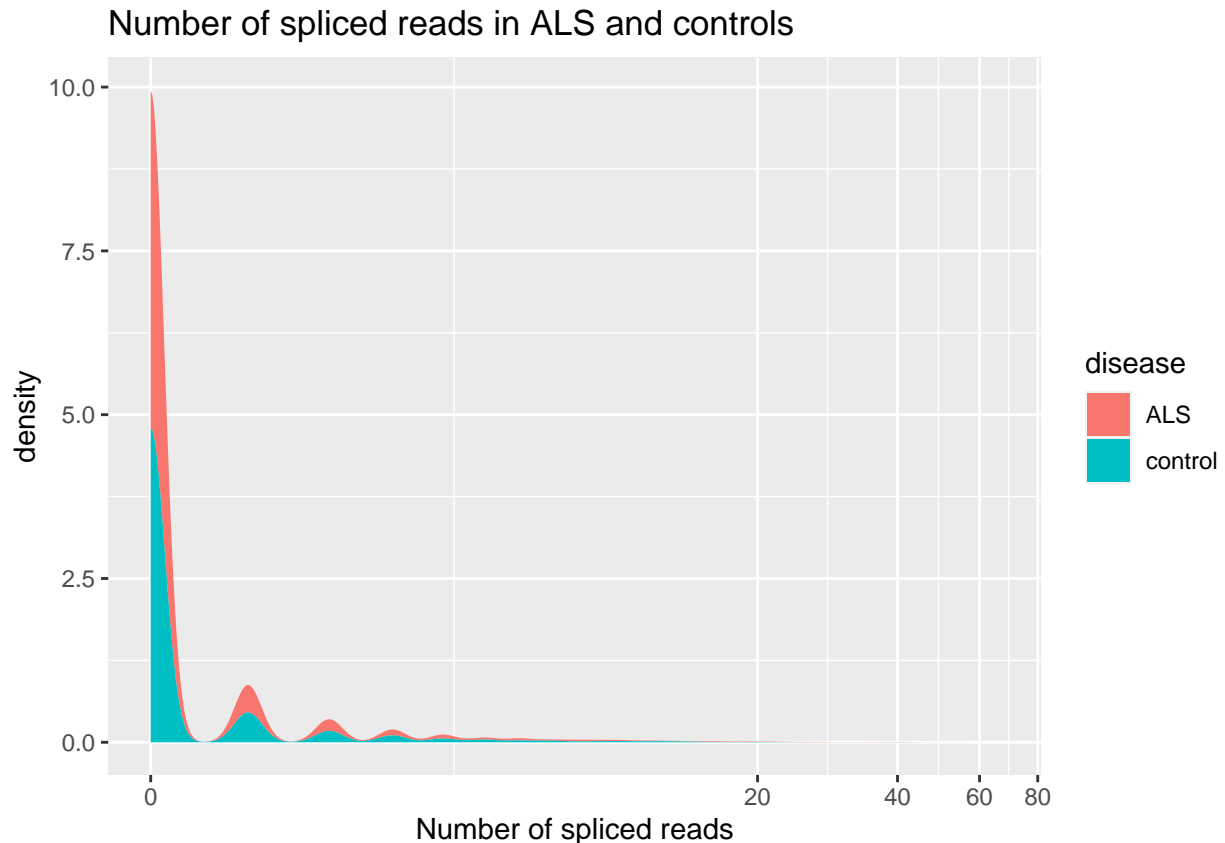
Figure 1: There is a greater number of spliced reads in ALS LCM neurons compared to control LCM neurons.

There are more spliced reads in the ALS cohort, with some more extreme values (Figure 1).

A new column was added to include information on the average count of each junction in ALS patients and control samples. This column was then split into two columns: ALS and control.

```
mean_per_junction <- group_column |>
  group_by(junction_name,disease) |>
  mutate(mean_n_spliced_reads = mean(n_spliced_reads)) |>
  ungroup() |>
  select(junction_name,disease,mean_n_spliced_reads) |>
  unique() |>
  pivot_wider(names_from = 'disease',
              values_from = 'mean_n_spliced_reads')
```

Wilcoxon test

The data were nested by junction name and a wilcoxon test was conducted to see if any junctions were differentially expressed in the ALS and control samples. The junctions were filtered for only those that were significant (i.e., wilcoxon test p-value < 0.05).

```
wilcox_tested_nested = group_column_nested |>
  mutate(wc_res = map(data,~{broom::tidy(wilcox.test(
    .x$n_spliced_reads ~ .x$disease, exact=FALSE))})) |>
```

```
  unnest(wc_res) |>
  arrange(p.value)

significant_junctions <- wilcox_tested_nested |>
  filter(p.value < 0.05)
```

33 junctions were significantly expressed.

Significantly expressed junctions

These significant junctions were grouped according to their disease group ("ALS" or "control") and the percentage of junctions more highly expressed in the ALS and control samples was calculated (Table 1).

Table 1: A greater proportion of junctions are more highly expressed in controls.

| higher_value	n	percent
ALS	8	0.2424242
control	25	0.7575758

76% of the significant junctions were more highly expressed in the controls, whereas only 24% were more highly expressed in ALS patients (Table 1). This was illustrated in a density plot (Figure 2).

```
sig_reads_groups <- group_column |>
  semi_join(significant_junctions, by=("junction_name"))

ggplot(sig_reads_groups, aes(x = n_spliced_reads, color = disease, fill = disease)) +
  geom_density(alpha = 0.5) +
  scale_x_continuous(trans = scales::pseudo_log_trans()) +
  labs(
    title = "Number of significant spliced reads in ALS and controls",
    x = "Number of spliced reads",
  )
```
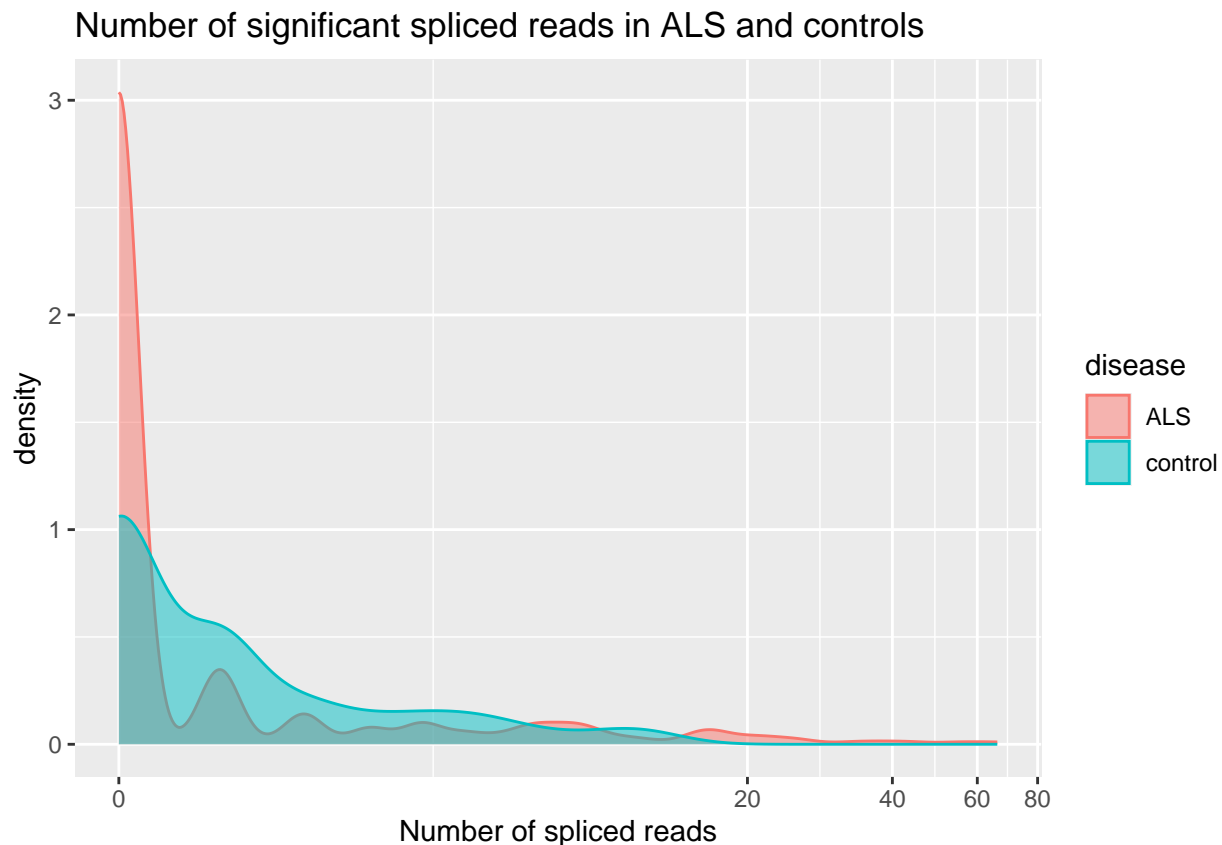
Figure 2: A greater number of spliced reads have significant expression in ALS LCM neurons compared to control LCM neurons.

Cryptic expression of different junctions in ALS and control samples

```
mean_per_junction |>
  semi_join(significant_junctions, by=("junction_name")) |>
  mutate(higher_value = ifelse(ALS>control,"ALS","control")) |>
  separate(junction_name, sep = '\\|',into = c("gene","junc_cat","n_datasets_junction_found"),convert =
  ggplot(aes(x = junc_cat, fill = higher_value)) +
  geom_bar() +
  labs(
    title = "Most of the junctions expressed higher in controls are annotated events",
    x = "Junction category",
    y = "Number of junctions",
    fill = "Group",
  )
```
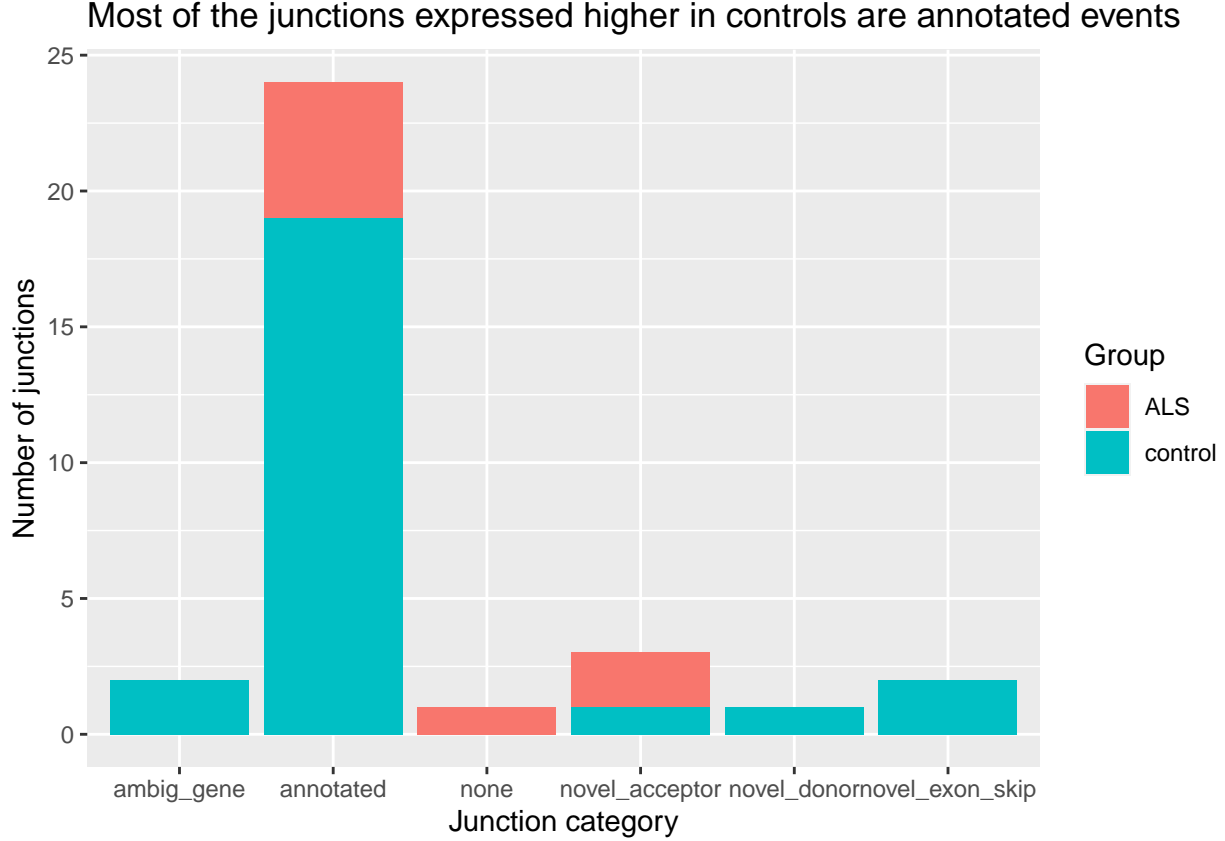
Figure 3: Annotated events are expressed significantly higher in control LCM neurons compared to other types of junction events. An *ambig_gene* junction overlaps exons originating from more than one gene. An *annotated* junction matches an existing intron. *None* means the junction has neither end overlapping a known exon. One end of a *novel_acceptor* or *novel_donor* junction matches that of a known exon. Both ends of a *novel_exon_skip* junction overlap a known exon boundary.

Datasets supporting cryptic and annotated STMN2 events

Next, we investigated the number of datasets (from knockdown studies) that have found significantly higher expression of these TDP43 events.

Table 2: Average number of datatsets the *STMN2* junction events are found in.

higher_value	mean_n_datasets
ALS	3.5
control	1.0

The annotated events found more highly expressed in control motor neurons were unique to a single TDP-43 knockdown, while the junction events in ALS motor neurons were more commonly found in studies (Table 2).

Investigating *STMN2* events

STMN2 events are more highly expressed in low accumulation of pTDP-43
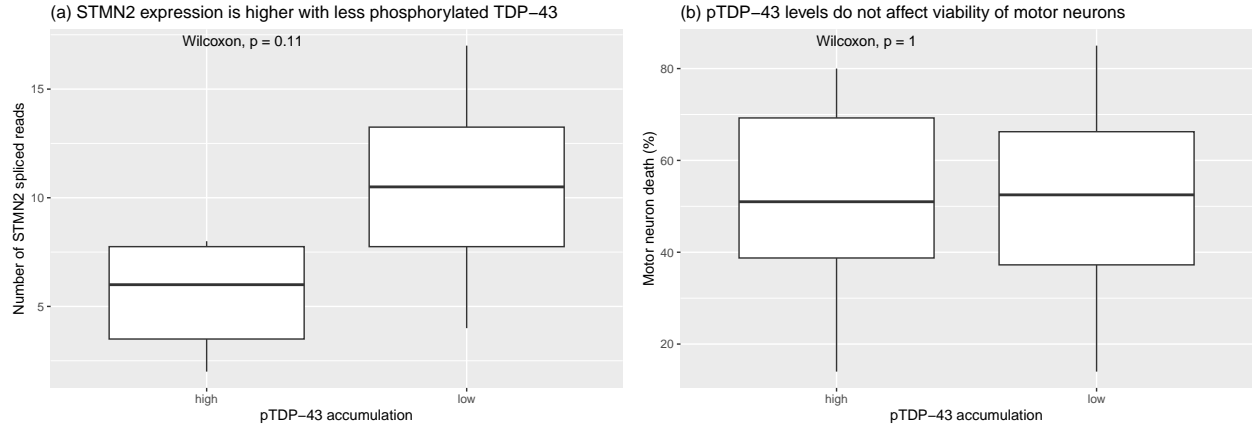
Figure 4: Effect of accumulation of phosphorylated TDP-43 (pTDP-43) on *STMN2* junction events. (a) There is no significant difference in *STMN2* expression between low and high pTDP-43 accumulation. (b) There is no significant difference in LCM neuron viability between low and high pTDP-43 accumulation.

Figure 4a shows greater expression of cryptic STMN2 junction events with lower levels of phosphorylated TDP-43 (pTDP-43), however we did not account for the normal expression levels of *STMN2* here - it could be that *STMN2* has low expression levels, which means we would see less cryptic splicing. Nonetheless, these results were not significant (p = 0.11). There was no effect (of pTDP-43 levels) on the viability of LCM neurons (Figure 4b).

These *STMN2* junction events were then grouped into "annotated" or "cryptic" events based on their start and end coordinates. The expression of annotated and cryptic *STMN2* junction events was analysed in LCM neurons from ALS and control samples, but also between low and high pTDP-43 accumulation.
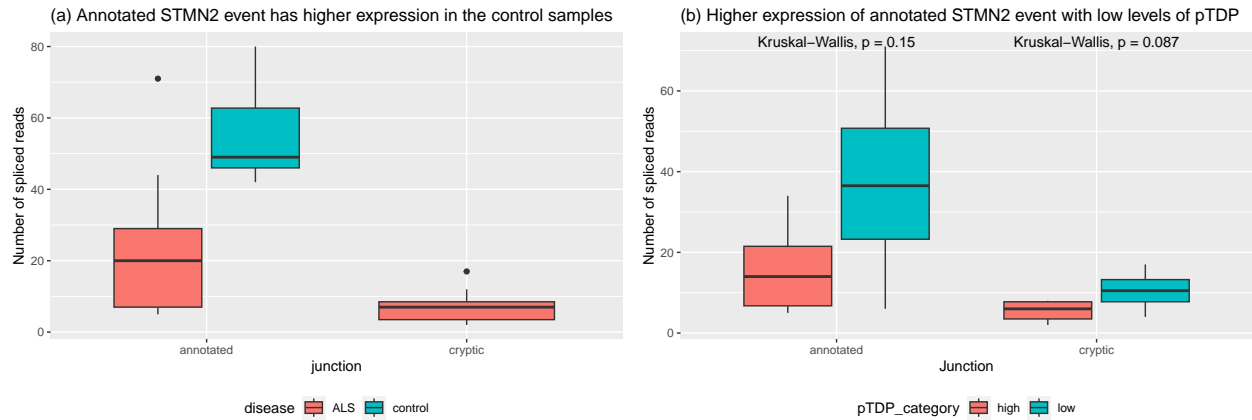

Figure 5: Expression levels of annotated and cryptic *STMN2* events between (a) ALS and control LCM neurons and (b) low and high pTDP-43 accumulation.

Consistent with data in Figure 3, there was greater expression of annotated events in control LCM neurons; whereas there was no difference in expression between annotated and cryptic events in ALS LCM neurons (Figure 5a). There was greater expression of annotated events in low levels of pTDP-43, as also shown in Figure 4a, however these results were not significant (Figure 5b).
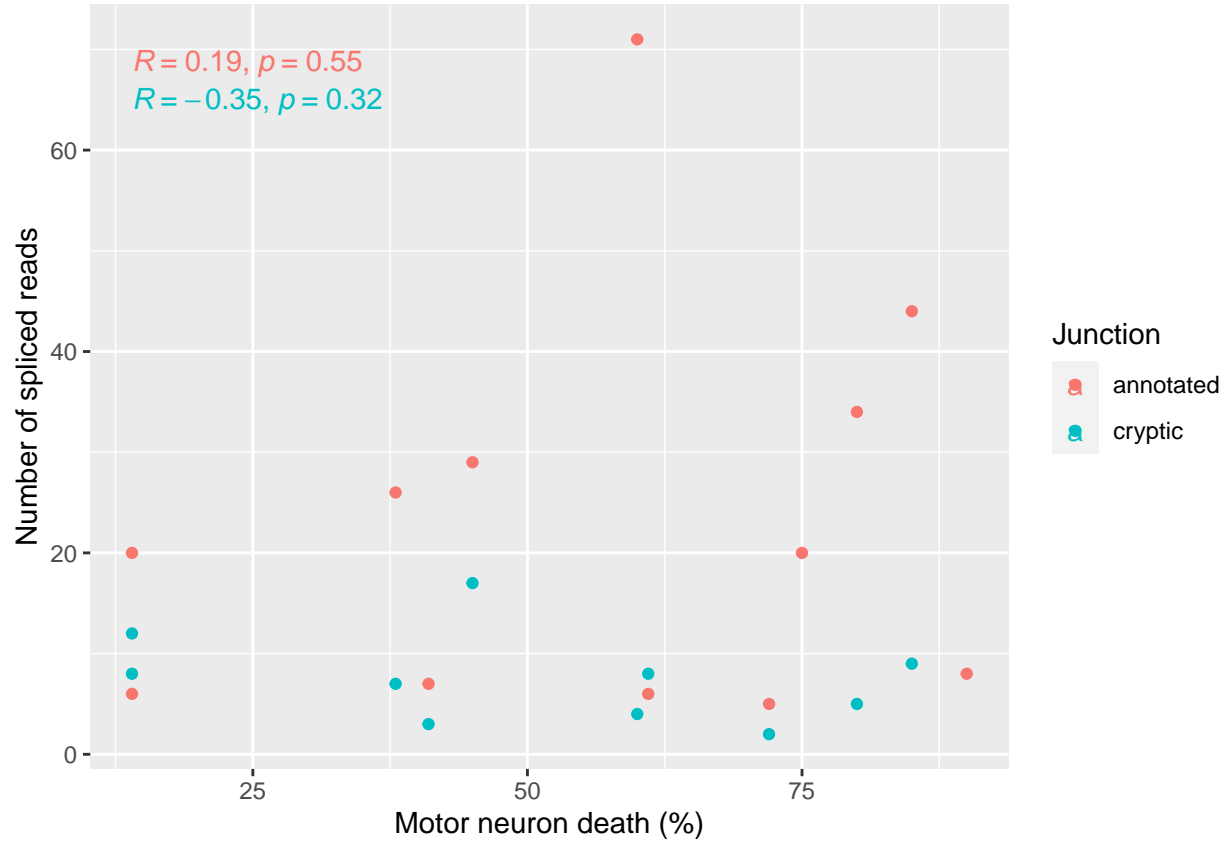
Figure 6: No correlation between motor neuron death and *STMN2* junction events.

```
psi_stmn2 <- psi_stmn2 |>
  group_by(sample_name) |>
  mutate(total_counts = annotated + cryptic) |>
  mutate(psi = cryptic / total_counts)
```
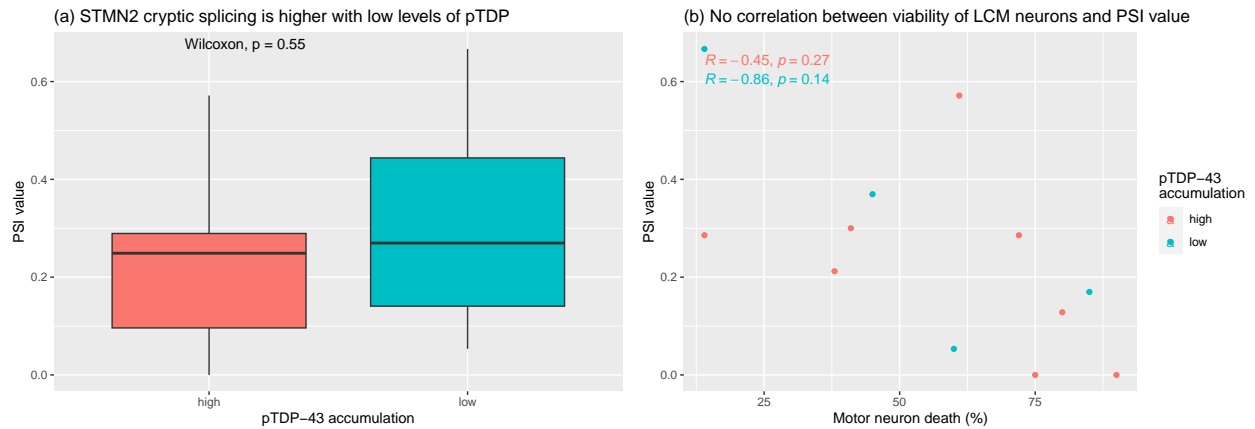

Figure 7: Effect of pTDP-43 accumulation on PSI value for *STMN2* cryptic splicing events and viability of LCM neurons. (a) There is no significant difference in PSI value between low and high pTDP-43 accumulation. (b) There is a weak negative correlation (not significant) between motor neuron death and PSI value, but no difference between low and high pTDP-43 levels.