# 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.

Number of spliced reads in ALS and controls
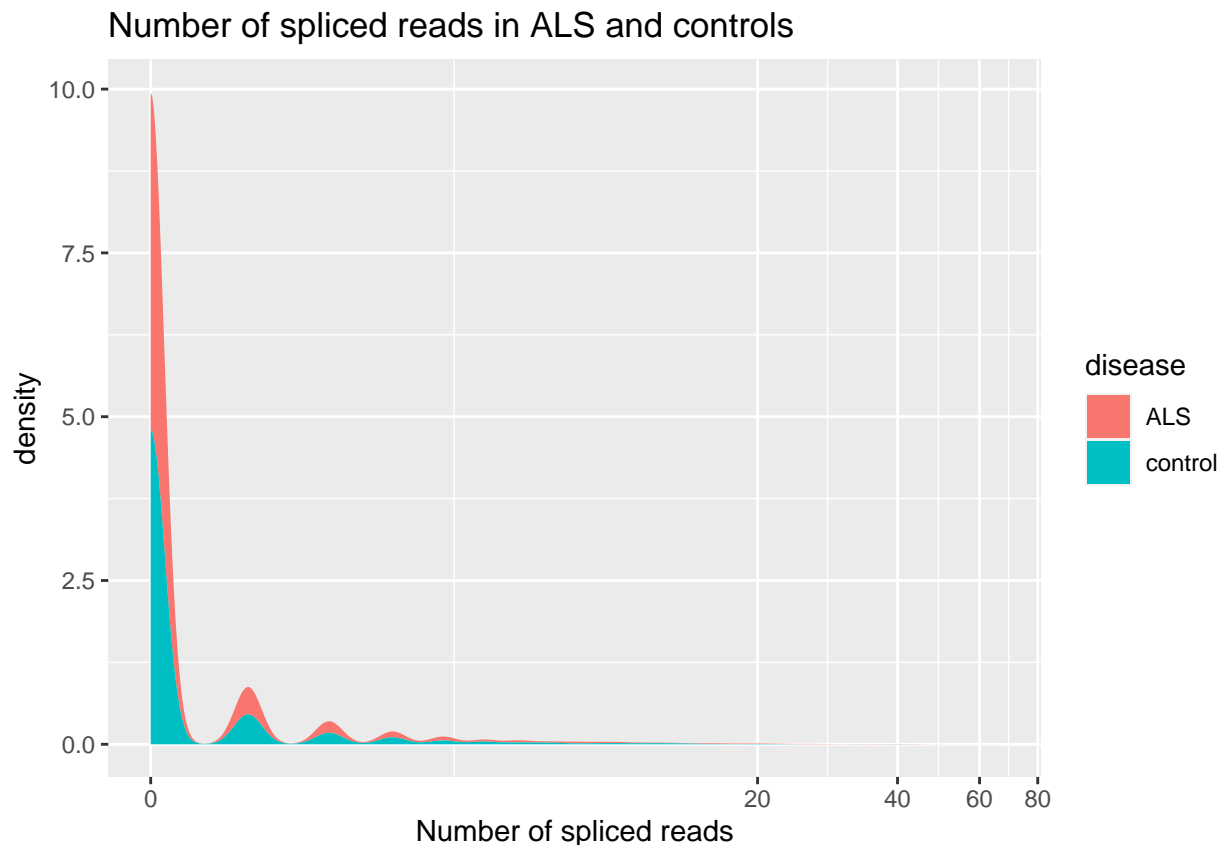
There are more spliced reads in the ALS cohort, with some more extreme values.

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: 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.

This was illustrated in a density plot:

```
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",
  )
```

Number of significant spliced reads in ALS and controls
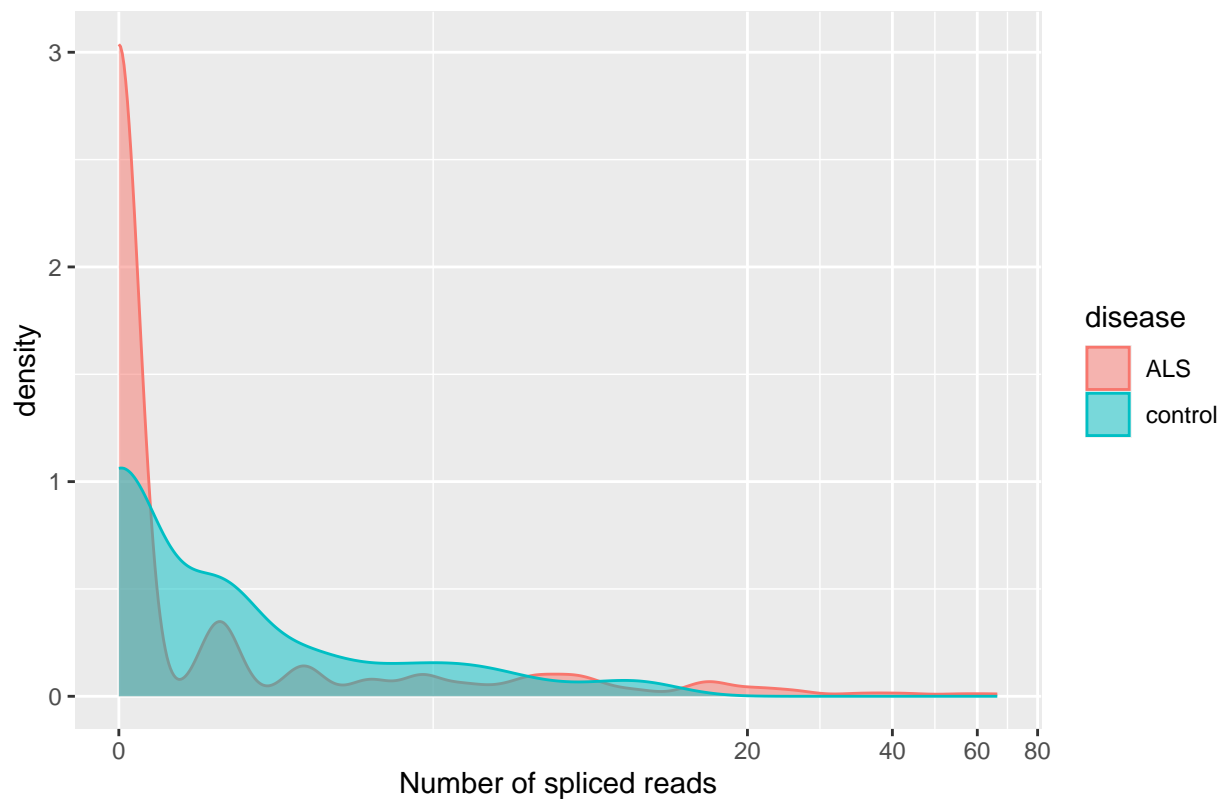
Cryptic expression of different junctions in ALS and control samples

```
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",
  )
```

Most of the junctions expressed higher in controls are annotated events
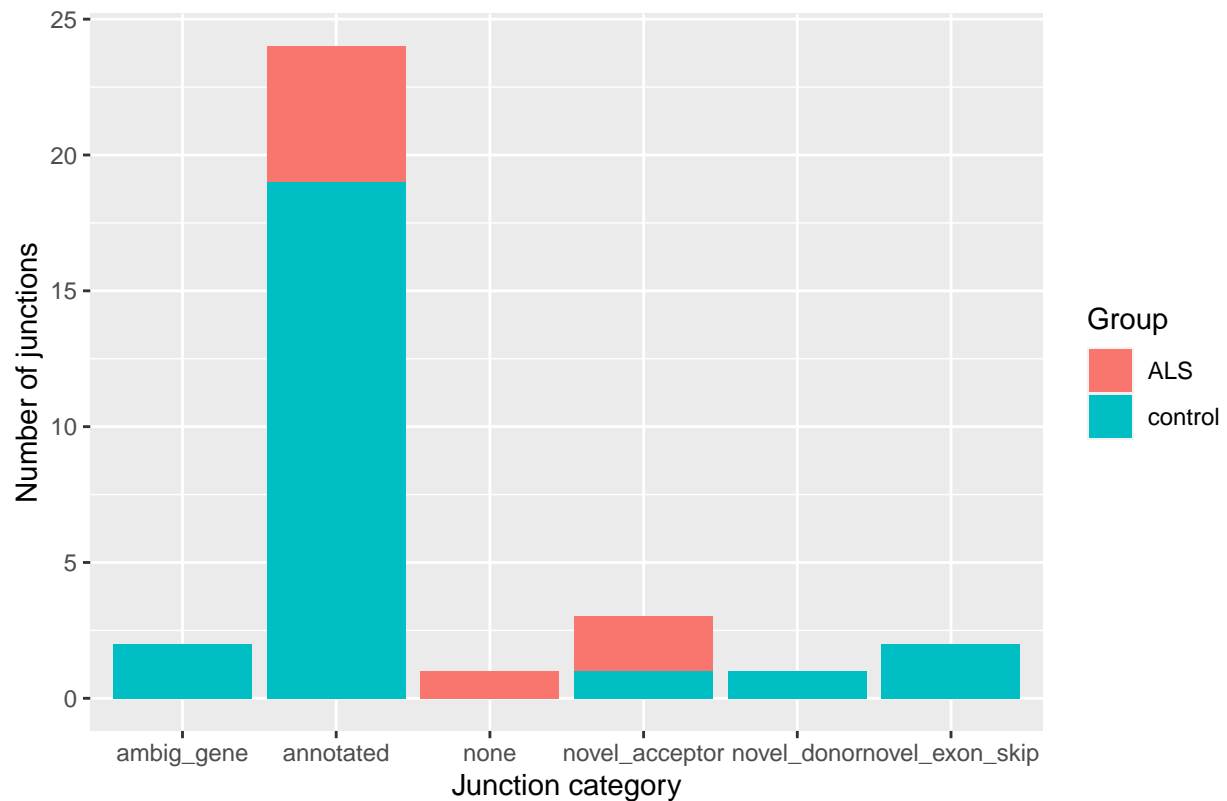
Mean_n_datasets_junction_found

| higher_value	mean_n_datasets
ALS	3.5
control	1.0

Reading in pTDP-43 data

```r
pTDP_vs_STMN2_reads_boxplot <- spliced_reads_pdtp_select |>
  filter(grepl("STMN2", junction_name)) |>
  drop_na() |>
  ggplot(aes(x = pTDP_category, y = n_spliced_reads)) +
  geom_boxplot() +
  theme(legend.position = "none") +
  labs(
    title = "STMN2 expression is higher with less phosphorylated TDP-43",
    x = "pTDP-43 accumulation",
    y = "Number of STMN2 spliced reads")
```

```
pTDP_vs_STMN2_reads_boxplot + stat_compare_means()

pTDP_vs_MNdeath_boxplot <- spliced_reads_pdtp_select |>
  filter(grepl("STMN2", junction_name)) |>
  drop_na() |>
  ggplot(aes(x = pTDP_category, y = MN_death)) +
  geom_boxplot() +
  theme(legend.position = "none") +
  labs(
    title = "pTDP-43 levels do not affect viability of motor neurons",
    x = "pTDP-43 accumulation",
    y = "Motor neuron death (%)")

pTDP_vs_MNdeath_boxplot + stat_compare_means()
```
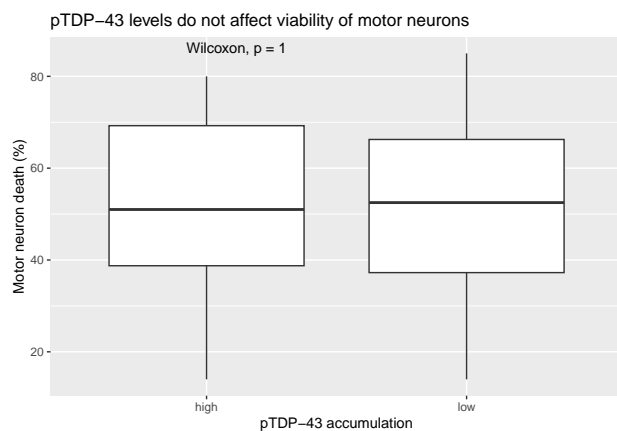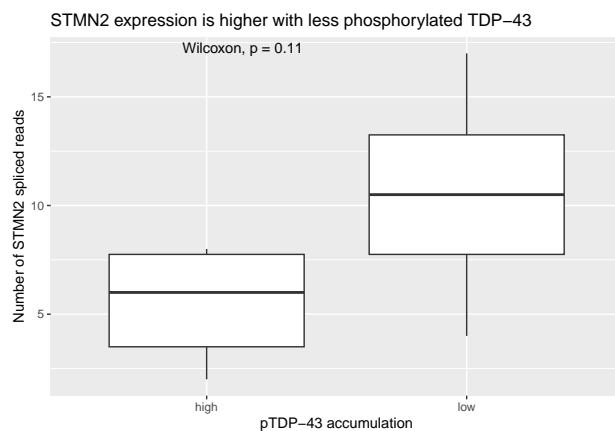

STMN2 data

```
STMN2_reads_vs_disease_boxplot <- spliced_reads_pdtp_stmn2 |>
  ggplot(aes(x = junction, y = n_spliced_reads, fill = disease)) +
  geom_boxplot() +
  labs(
    title = "Annotated STMN2 event has higher expression in the control samples",
    y = "Number of spliced reads"
  ) +
  theme(legend.position = "bottom")

STMN2_reads_vs_disease_boxplot + stat_compare_means()
```

```
## Warning: Computation failed in 'stat_compare_means()'
## Caused by error in 'mutate()':
## ! Problem while computing 'p = purrr::map(...)'.
## Caused by error in 'purrr::map()':
## i In index: 2.
## i With name: x.2.
## Caused by error in 'kruskal.test.default()':
## ! all observations are in the same group
```

```
STMN2_reads_vs_pTDPcat_boxplot <- spliced_reads_pdtp_stmn2 |>
  drop_na() |>
  ggplot(aes(x = junction, y = n_spliced_reads, fill = pTDP_category)) +
  geom_boxplot() +
  labs(
    title = "Higher expression of annotated STMN2 event with low levels of pTDP",
    x = "Junction",
    y = "Number of spliced reads"
  ) +
  theme(legend.position = "bottom")

STMN2_reads_vs_pTDPcat_boxplot + stat_compare_means()
```
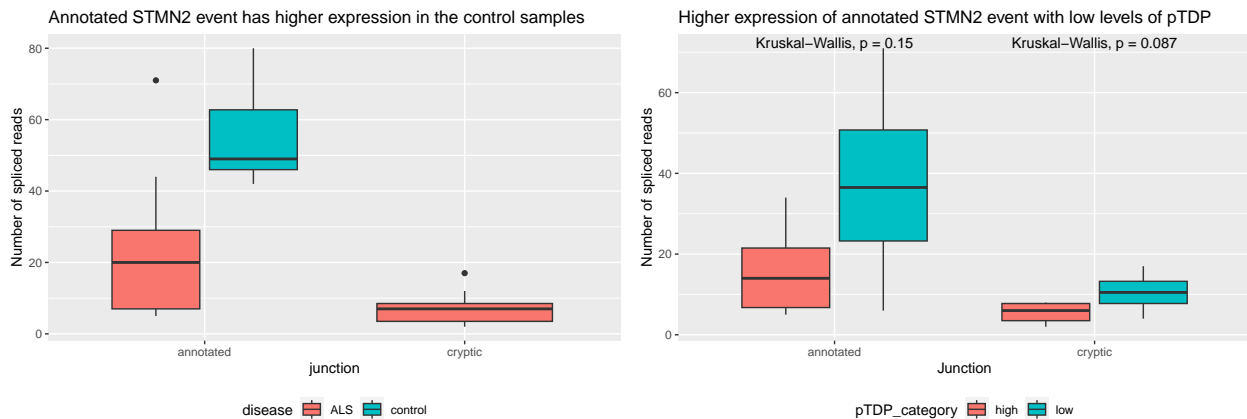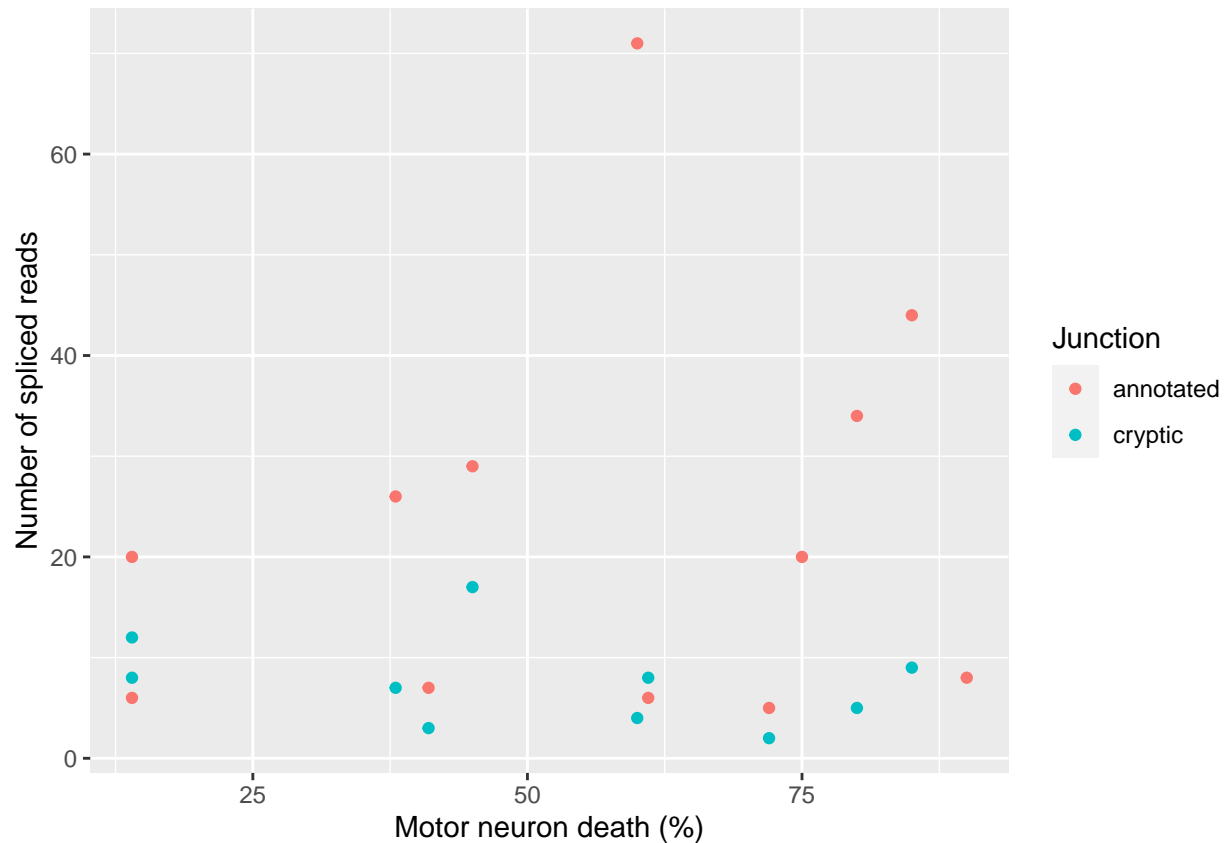


```
spliced_reads_pdtp_stmn2 |>
  drop_na() |>
  ggplot(aes(x = MN_death, y = n_spliced_reads, color = junction)) +
  geom_point() +
  labs(
    x = "Motor neuron death (%)",
    y = "Number of spliced reads",
    color = "Junction"
  )
```

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

```
psi_boxplot1 <- psi_stmn2 |>
  drop_na() |>
  ggplot(aes(x = pTDP_category, y = psi, fill = pTDP_category)) +
  geom_boxplot() +
  theme(legend.position = "none") +
  labs(
    title = "STMN2 cryptic splicing is higher with low levels of pTDP",
    x = "pTDP-43 accumulation",
    y = "PSI value"
  )

psi_boxplot1 + stat_compare_means()

psi_boxplot2 <- psi_stmn2 |>
  drop_na() |>
  ggplot(aes(x = MN_death, y = psi, color = pTDP_category)) +
  geom_point() +
  labs(
    x = "Motor neuron death (%)",
    y = "PSI value",
```

```
    color = "pTDP-43 \naccumulation"
  )
```

```
psi_boxplot2 + stat_cor()
```


STMN2 cryptic splicing is higher with low levels of pTDP