Clinical, gene and mutation data of cancers with cryptic STMN2 events

Cancers with STMN2 expression

```
STMN2_clinical |>
  ggplot(aes(x = fct_rev(fct_infreq(cancer_type)))) +
  geom_bar(aes(fill = cancer_type)) +
  coord_flip() +
  labs(
    x = "Cancer Type",
    y = "Number of Cases",
    title = "STMN2 is expressed in mostly breast and brain cancers"
  ) +
  theme(legend.position = "none", plot.title = element_text(size=10))
```
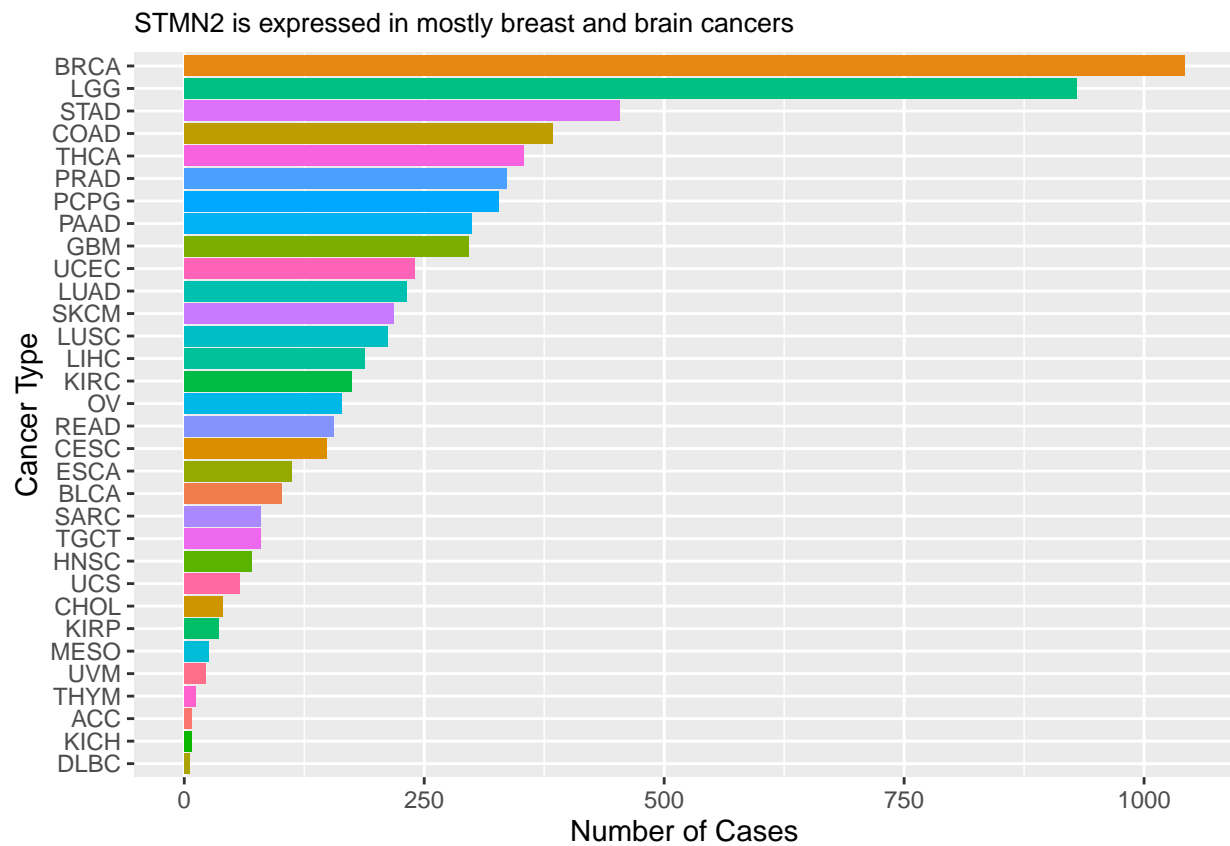

Figure 1: STMN2 events are found mostly in breast and brain cancer patients. BRCA = breast cancer; LGG = low-grade gliomas (brain tumours)

Cancers with cryptic STMN2 expression

```
STMN2_clinical_jir_cryptic <- STMN2_clinical_jir |>
  filter(STMN2_cryptic_coverage > 2)
```

Primary sites of cancers

```
STMN2_clinical_jir_cryptic |>
  drop_na() |>
  ggplot(aes(x = fct_rev(fct_infreq(gdc_primary_site)))) +
  geom_bar(aes(fill = gdc_primary_site)) +
  coord_flip() +
  labs(
    x = "Primary Site of Cancer",
    y = "Number of Cases",
    title =
  "Primary sites of cancers with cryptic STMN2 events are mostly the adrenal gland and brain"
  ) +
  theme(legend.position = "none", plot.title = element_text(size=9))
```
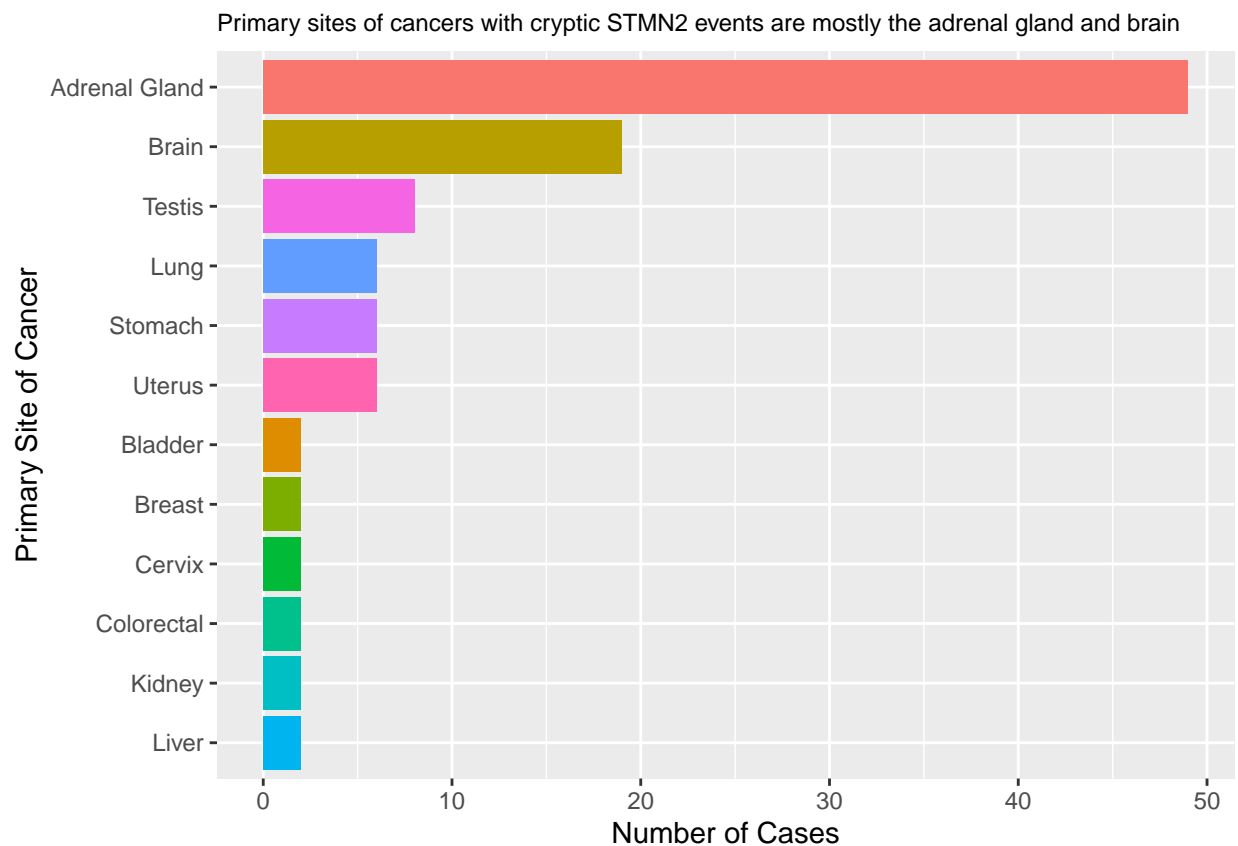

Figure 2: Cryptic STMN2 events are found mostly in the adrenal gland and brain.

Interestingly, Figure 2 shows that cryptic STMN2 expression is low in the breast. Comparing this to Figure

1, this shows that the STMN2 events expressed in breast cancers may be mostly annotated non-cryptic events.

Cryptic STMN2 junction coverage in different sites of cancers

```
STMN2_clinical_jir_cryptic |>
  drop_na() |>
  filter(cgc_primary_site != "") |>
  ggplot(aes(x = junction_avg_coverage, y = fct_reorder(gdc_primary_site,
                                            junction_avg_coverage, median))) +
  geom_boxplot(aes(fill = gdc_primary_site)) +
  labs(
    x = "Junction Average Coverage",
    y = "Primary Site of Cancer",
  ) +
  theme(legend.position = "none", plot.title = element_text(size=10)) +
  geom_signif(comparisons = list(c("Stomach", "Breast"), c("Stomach", "Brain")),
              map_signif_level = TRUE,
              y_position = c(75, 80))
```
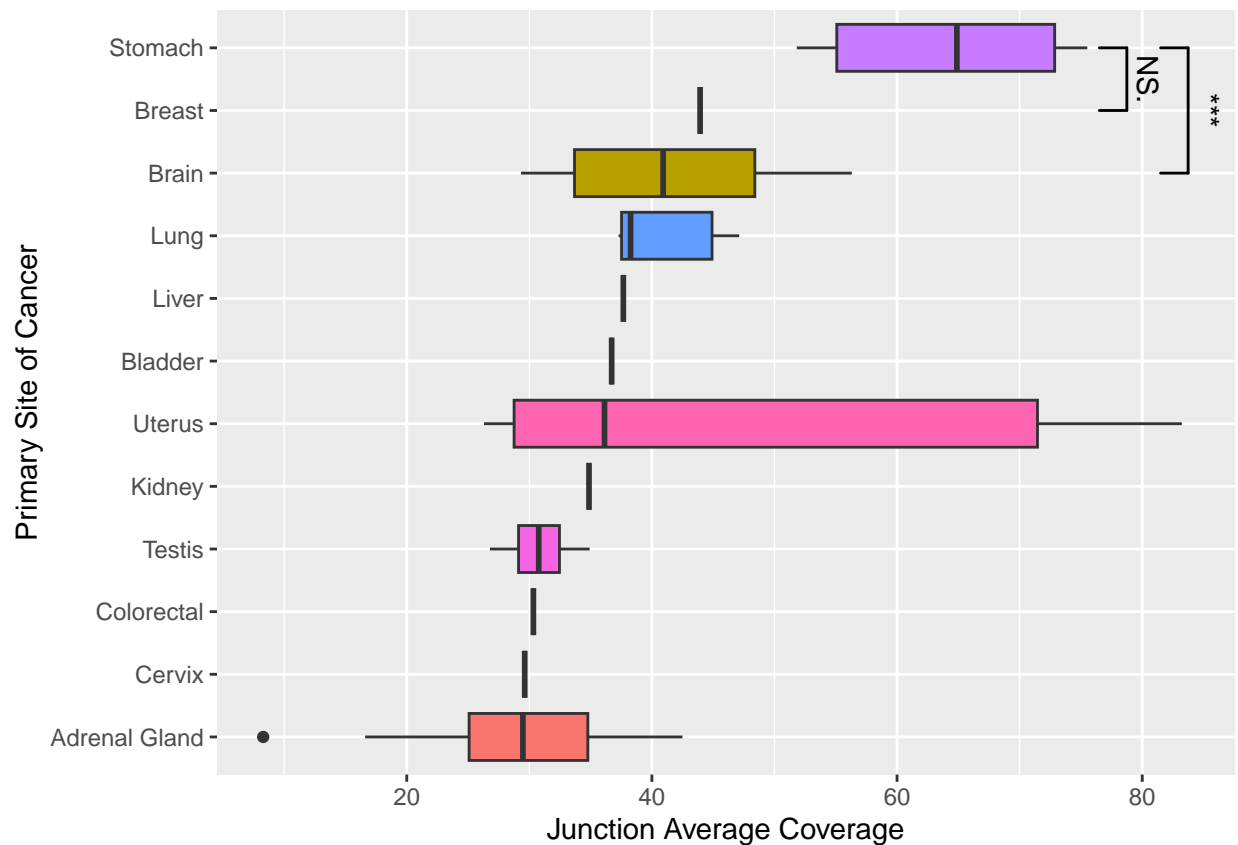

Figure 3: Stomach and breast cancers are the most deeply sequenced.

```
STMN2_clinical_jir_cryptic |>
  drop_na() |>
```

```
filter(cgc_primary_site != "") |>
ggplot(aes(x = STMN2_cryptic_coverage,
           y = fct_reorder(gdc_primary_site, STMN2_cryptic_coverage, median))) +
geom_boxplot(aes(fill = gdc_primary_site)) +
labs(
  x = "STMN2 Cryptic Coverage",
  y = "Primary Site of Cancer",
) +
theme(legend.position = "none", plot.title = element_text(size=10))
```
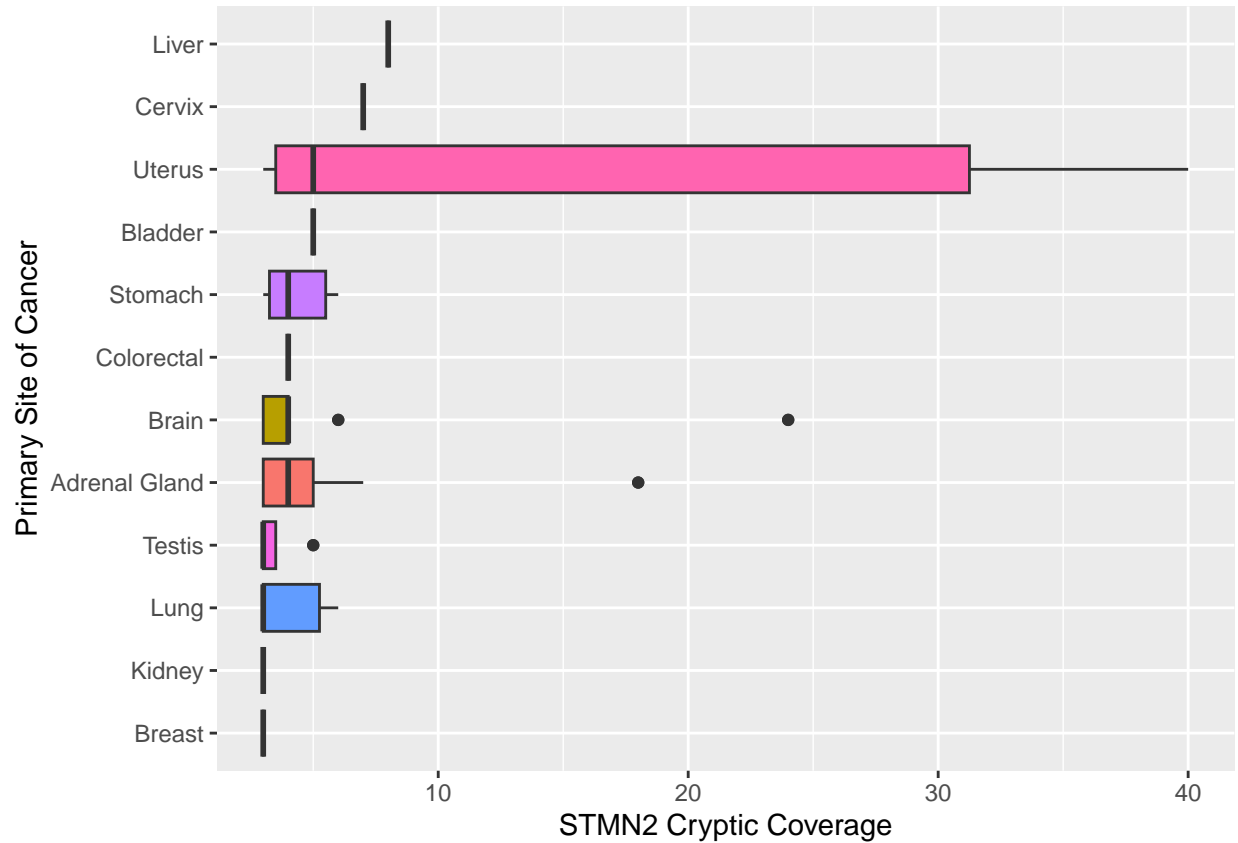

Figure 4: On average, cancers in the stomach and breast have the greatest number of reads supporting cryptic STMN2 events. Cancers of the brain have significantly fewer reads surpporting cryptic events.

Which cancers have the most cryptic STMN2 events?

Table 1: Breast and brain cancers have high cryptic STMN2 expression.

cancer_type	n	percent
Pheochromocytoma and Paraganglioma	49	0.4622642
Glioblastoma Multiforme	15	0.1415094
Testicular Germ Cell Tumors	8	0.0754717
Stomach Adenocarcinoma	6	0.0566038
Brain Lower Grade Glioma	4	0.0377358

cancer_type	n	percent
Lung Squamous Cell Carcinoma	4	0.0377358
Uterine Corpus Endometrial Carcinoma	4	0.0377358
Bladder Urothelial Carcinoma	2	0.0188679
Breast Invasive Carcinoma	2	0.0188679
Cervical Squamous Cell Carcinoma and Endocervical Adenocarcinoma	2	0.0188679
Kidney Renal Clear Cell Carcinoma	2	0.0188679
Liver Hepatocellular Carcinoma	2	0.0188679
Lung Adenocarcinoma	2	0.0188679
Rectum Adenocarcinoma	2	0.0188679
Uterine Carcinosarcoma	2	0.0188679

Where are the cancers with cryptic STMN2 events located?

Table 2: Cancers with cryptic STMN2 events are found primarily in the adrenal gland and brain

gdc_primary_site	n	percent
Adrenal Gland	49	0.4622642
Brain	19	0.1792453
Testis	8	0.0754717
Lung	6	0.0566038
Stomach	6	0.0566038
Uterus	6	0.0566038
Bladder	2	0.0188679
Breast	2	0.0188679
Cervix	2	0.0188679
Colorectal	2	0.0188679
Kidney	2	0.0188679
Liver	2	0.0188679

TCGA Clinical Data

Mutation Counts

```
STMN2_cryptic_cBio_mutations |>
  ggplot(aes(x = fct_reorder(cancer_abbrev, mean_mutation_count, median), y = mean_mutation_count)) +
  geom_bar(stat = 'identity', aes(fill = cancer_abbrev)) +
  labs(
    x = "Cancer Type",
    y = "Mean Mutation Count",
    title = "Glioblastoma Multiforme has the greatest number of mutations"
  ) +
  theme(legend.position = "none")
```

Glioblastoma Multiforme has the greatest number of mutations
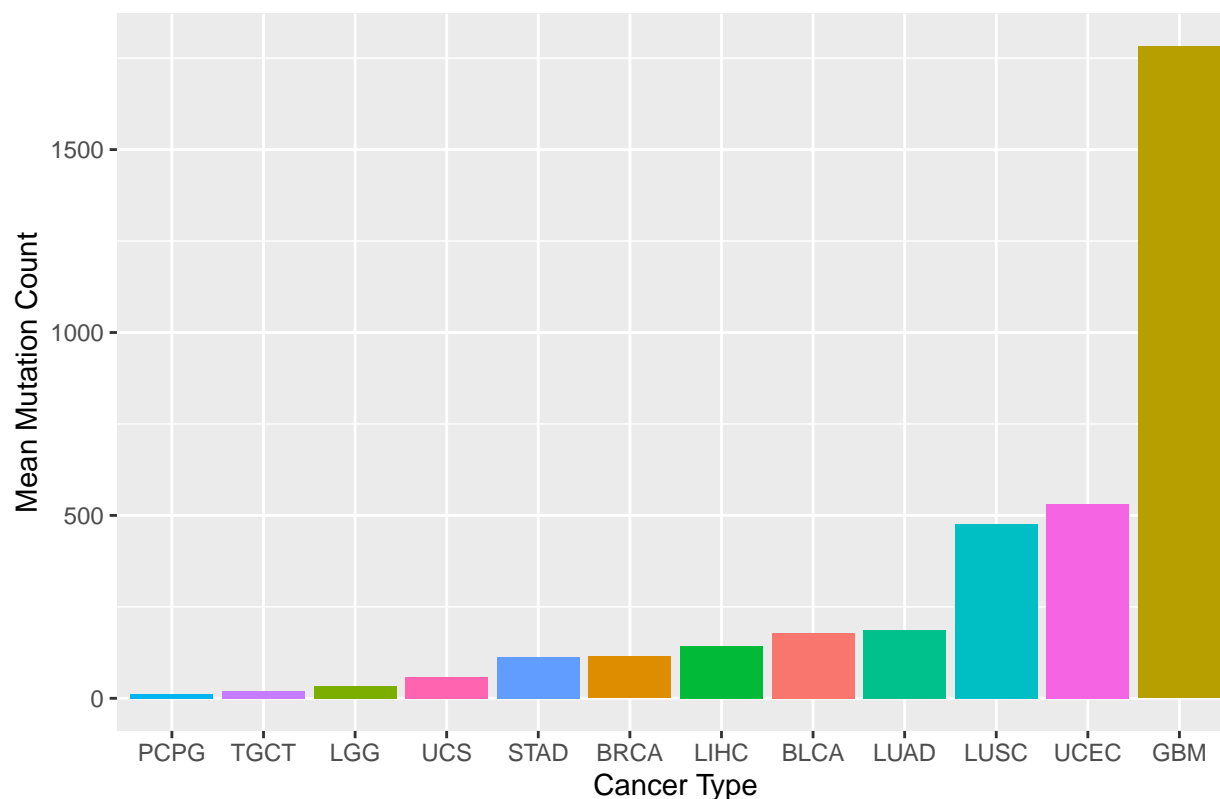
There is an outlier in GBM for mutation count, which heavily skews the mean for this cancer type.

```
STMN2_cryptic_cBio <- STMN2_cryptic_cBio |>
  mutate_at("mutation_count", as.numeric)
```

```
## Warning: There was 1 warning in `mutate()`.
## i In argument: `mutation_count = .Primitive("as.double")(mutation_count)`.
## Caused by warning:
## ! NAs introduced by coercion
```

```
STMN2_cryptic_cBio |>
  drop_na() |>
  filter(mutation_count < 2500) |>
  ggplot(aes(x = fct_reorder(cancer_abbrev, mutation_count, median), y = mutation_count)) +
  geom_boxplot() +
  labs(
    x = "Cancer Type",
    y = "Mutation Count",
    title = "UCEC and LUSC cancers have the greatest number of mutations"
  ) +
  geom_signif(comparisons = list(c("LUSC", "UCEC")),
map_signif_level = TRUE,
y_position = c(1000))
```

```
## Warning in wilcox.test.default(c(268, 268, 684, 684), c(60, 60, 1002, 1002:
## cannot compute exact p-value with ties
```

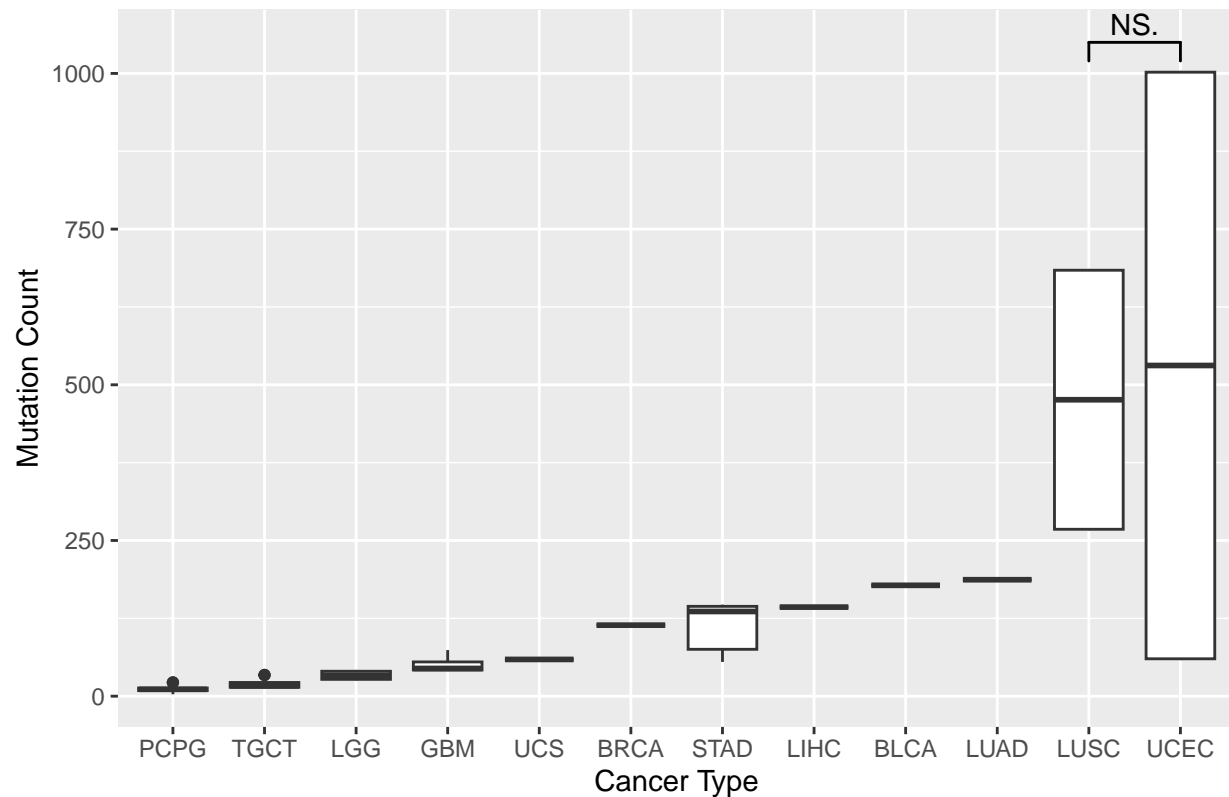**Figure 5**: Among cancer patients with STMN2 cryptic expression, uterine corpus endometrial carcinoma (UCEC) has the greatest number of mutations.

Mutation Data of one patient

```
patient_mutations |>
  count(Gene, sort = TRUE) |>
  filter(n > 17) |>
  ggplot(aes(x = fct_reorder(Gene, n, mean), y = n)) +
  geom_bar(stat = 'identity', aes(fill = Gene)) +
  labs(
    x = "Gene",
    y = "Mutation Count"
  ) +
  theme(legend.position = "none")
```
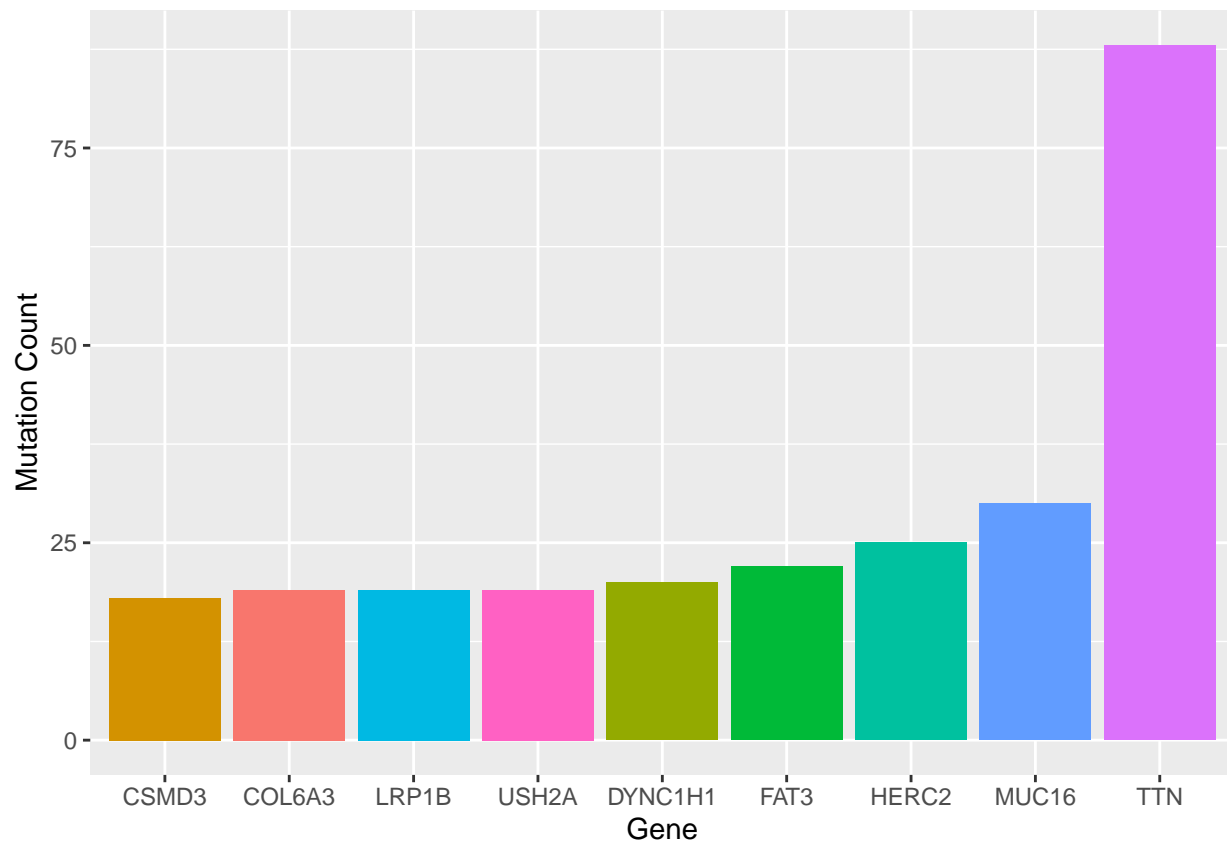
Figure 6: Among cancer patients with STMN2 cryptic expression, the titin gene has the greatest number of mutations.

```
patient_mutations |>
  filter(Gene == "TTN") |>
  n_distinct()
```

```
## [1] 88
```

```
patient_mutations |>
  filter(Gene == "TTN") |>
  unique() |>
  ggplot(aes(x = fct_rev(fct_infreq(Mutation.Type)))) +
  geom_bar(aes(fill = Mutation.Type)) +
  labs(
    x = "Mutation Type",
    y = "Count in TTN Gene"
  ) +
  theme(legend.position = "none")
```
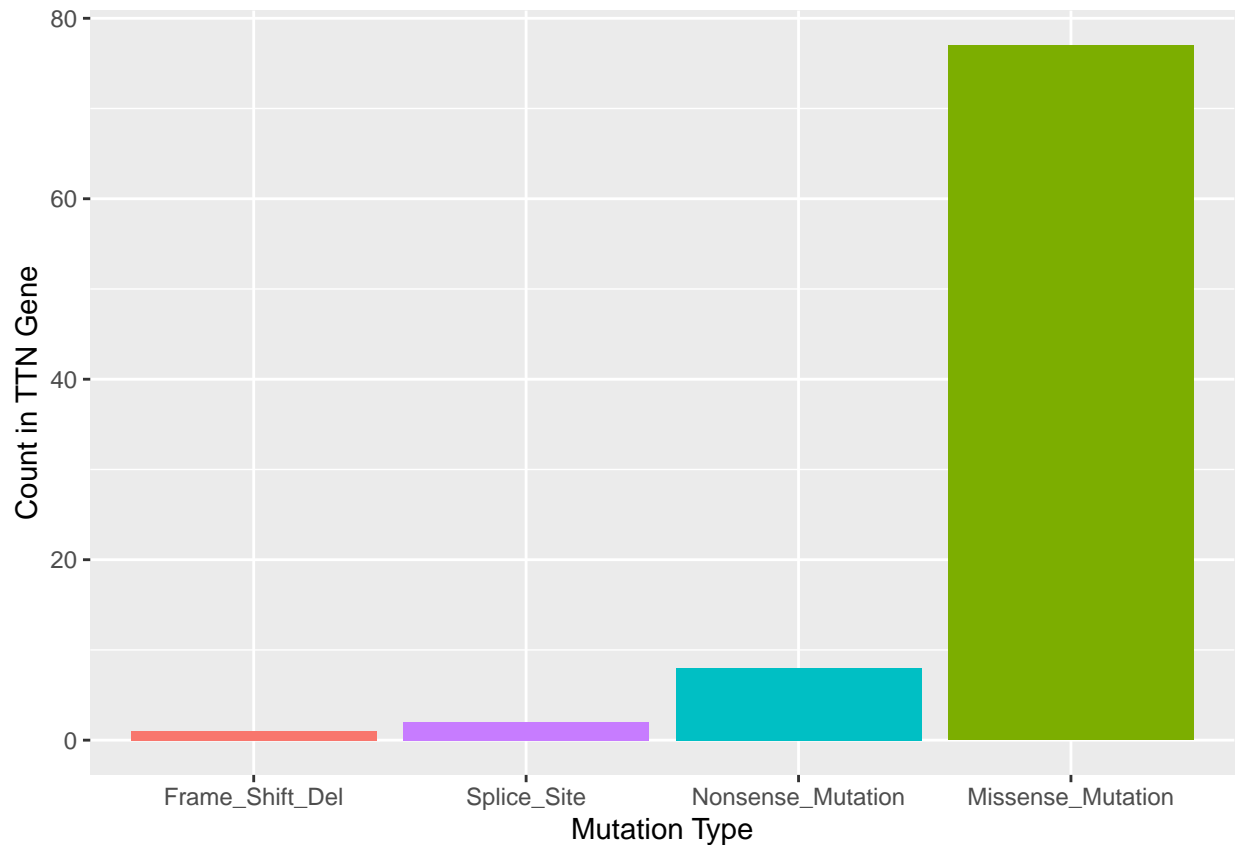
Figure 7: Most of the mutations seen in the titin gene are missense mutations.

STMN2 cryptic expression

```
STMN2_cryptic_cBio <- STMN2_cryptic_cBio |>
  mutate_at("mutation_count", as.numeric)

STMN2_cryptic_cBio |>
  drop_na() |>
  filter(mutation_count < 400, stmn2_cryptic_coverage > 1) |>
  ggplot(aes(x = as.factor(stmn2_cryptic_coverage), y = mutation_count)) +
  geom_boxplot() +
  labs(
    x = "Number of STMN2 cryptic events",
    y = "Mutation Count"
  ) +
  geom_signif(comparisons = list(c("6", "8"), c("5", "6"), c("3", "6")),
              map_signif_level = TRUE,
              y_position = c(250, 300, 200))
```
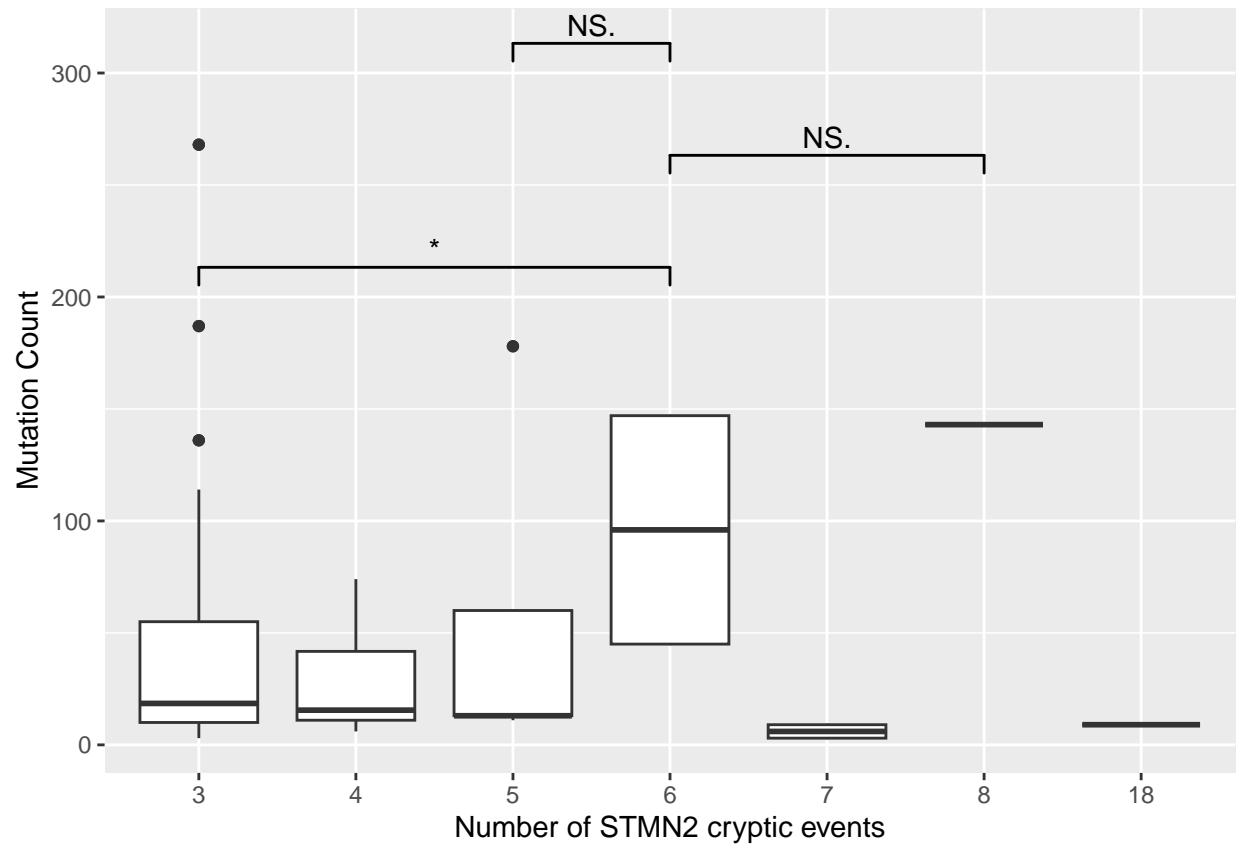
```
STMN2_cryptic_cBio |>
  drop_na() |>
  filter(mutation_count < 2500) |>
  ggplot(aes(x = stmn2_cryptic_coverage, y = mutation_count)) +
  geom_hex() +
  labs(
    x = "Number of STMN2 cryptic events",
    y = "Mutation Count"
  )
```
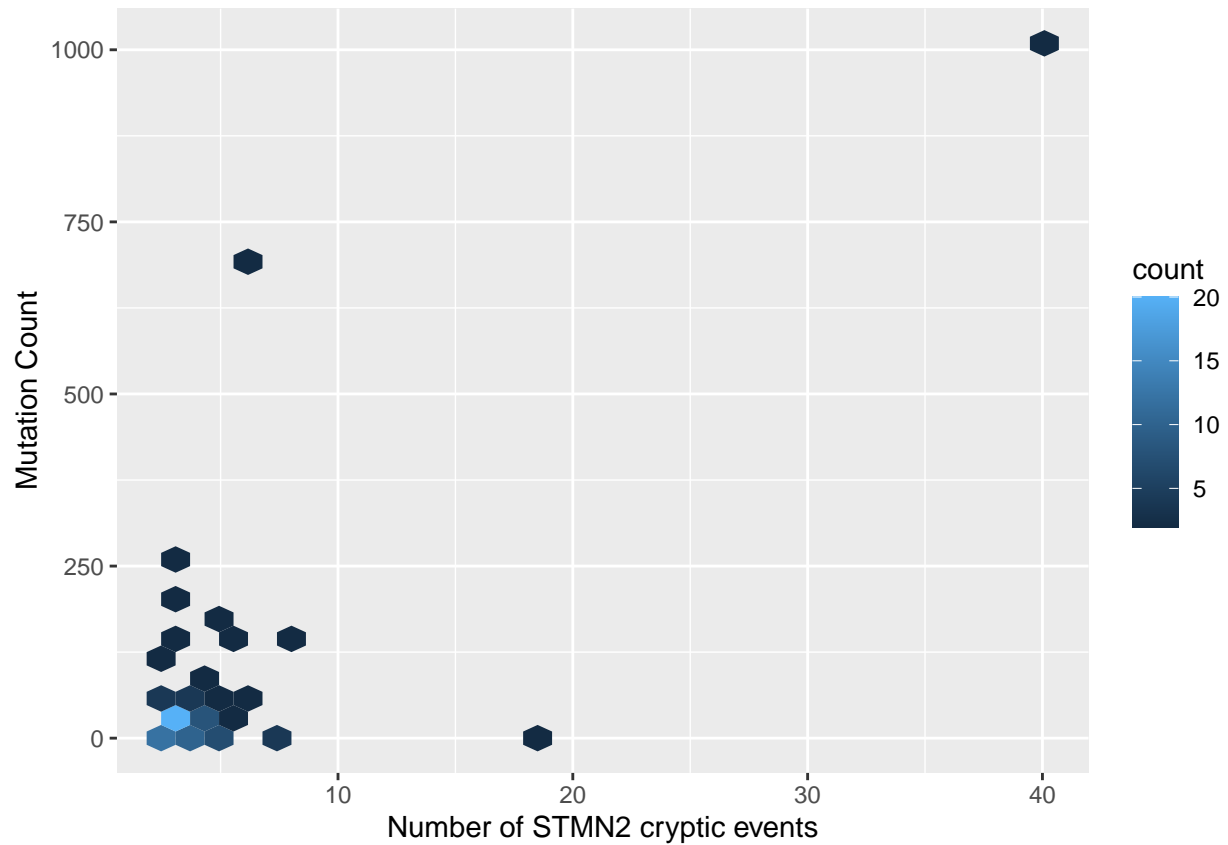
Figure 8: There is no correlation between number of STMN2 cryptic events and number of mutations.

Cancer Driver Genes

Mutations in each cancer gene

```
cosmic_patient_mutations |>
  count(Gene, sort = TRUE) |>
  filter(n > 11) |>
  ggplot(aes(x = fct_reorder(Gene, n, mean), y = n)) +
  geom_bar(stat = 'identity', aes(fill = Gene)) +
  labs(
    x = "Cancer Gene",
    y = "Mutation Count"
  ) +
  theme(legend.position = "none")
```
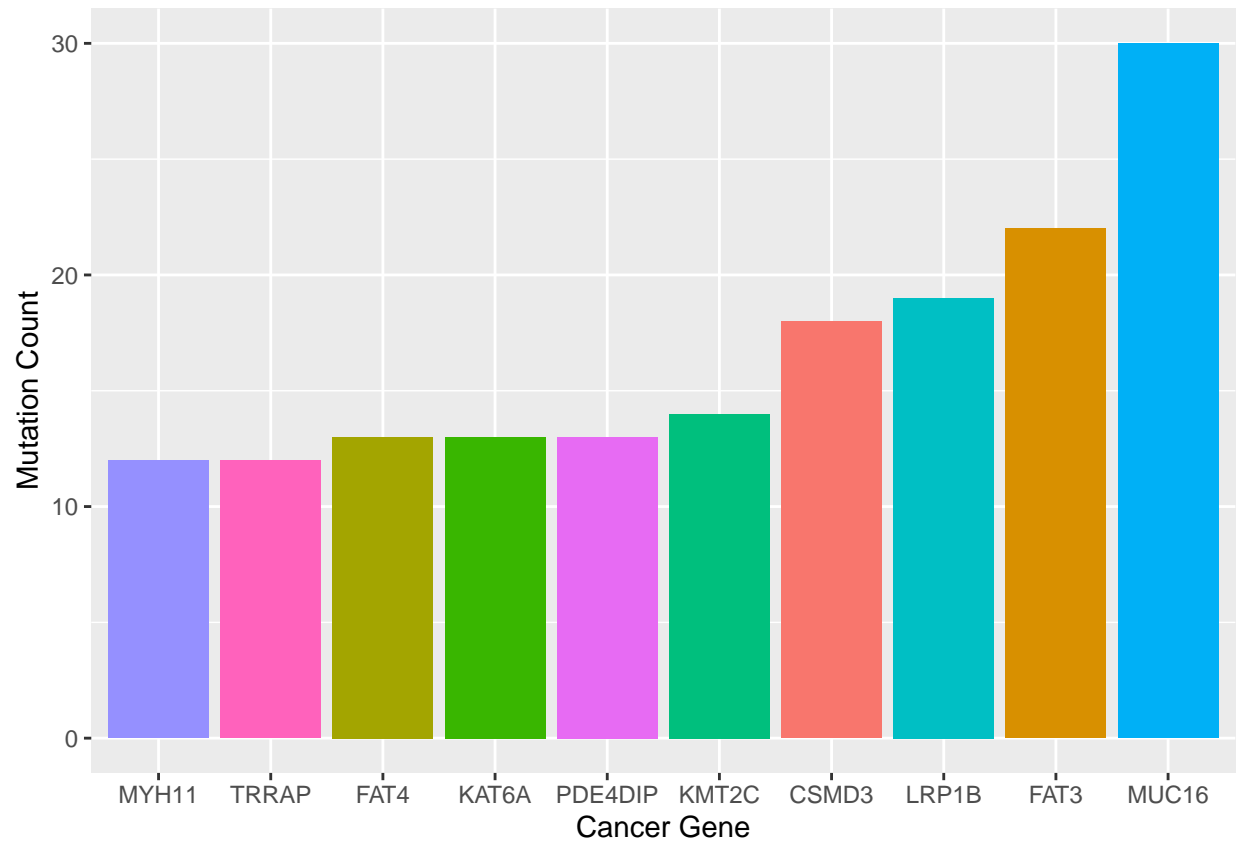
Figure 9: Ovarian cancer gene MUC16 has the greatest number of mutations.

```
cosmic_patient_mutations |>
  filter(Gene == "MUC16") |>
  ggplot(aes(x = fct_rev(fct_infreq(Mutation.Type)))) +
  geom_bar(aes(fill = Mutation.Type)) +
  labs(
    x = "Mutation Type",
    y = "Count in MUC16 Gene"
  ) +
  theme(legend.position = "none")
```
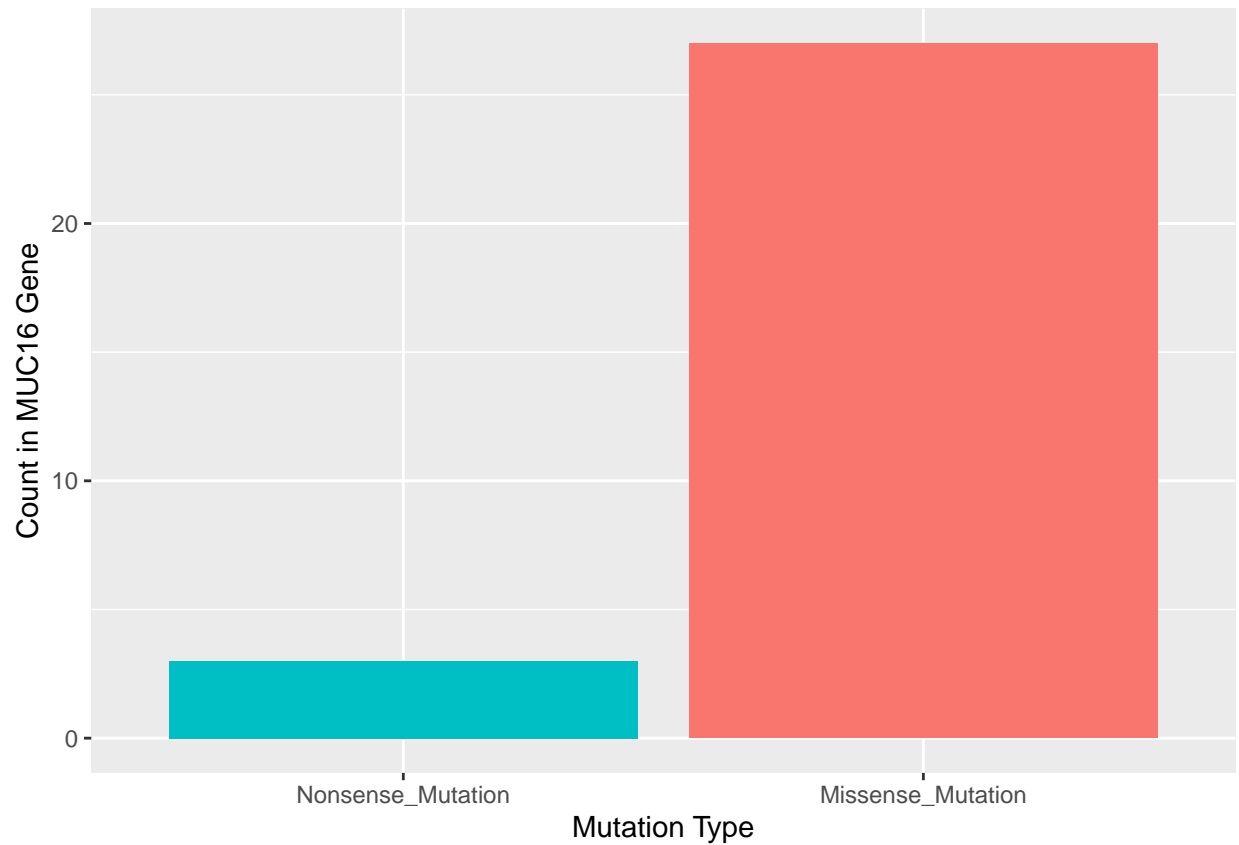
Figure 10: Ovarian cancer gene MUC16 has mostly missense mutations.

Types of cancer genes

```
cosmic_patient_mutations |>
  drop_na() |>
  ggplot(aes(x = role_in_cancer)) +
  geom_bar(aes(fill = role_in_cancer)) +
  labs(
    x = "Role of the gene in cancer",
    y = "Number of mutations",
    title = "Most of the mutations are found in tumour suppressor genes"
  ) +
  theme(legend.position = "none", plot.title = element_text(size=12))
```
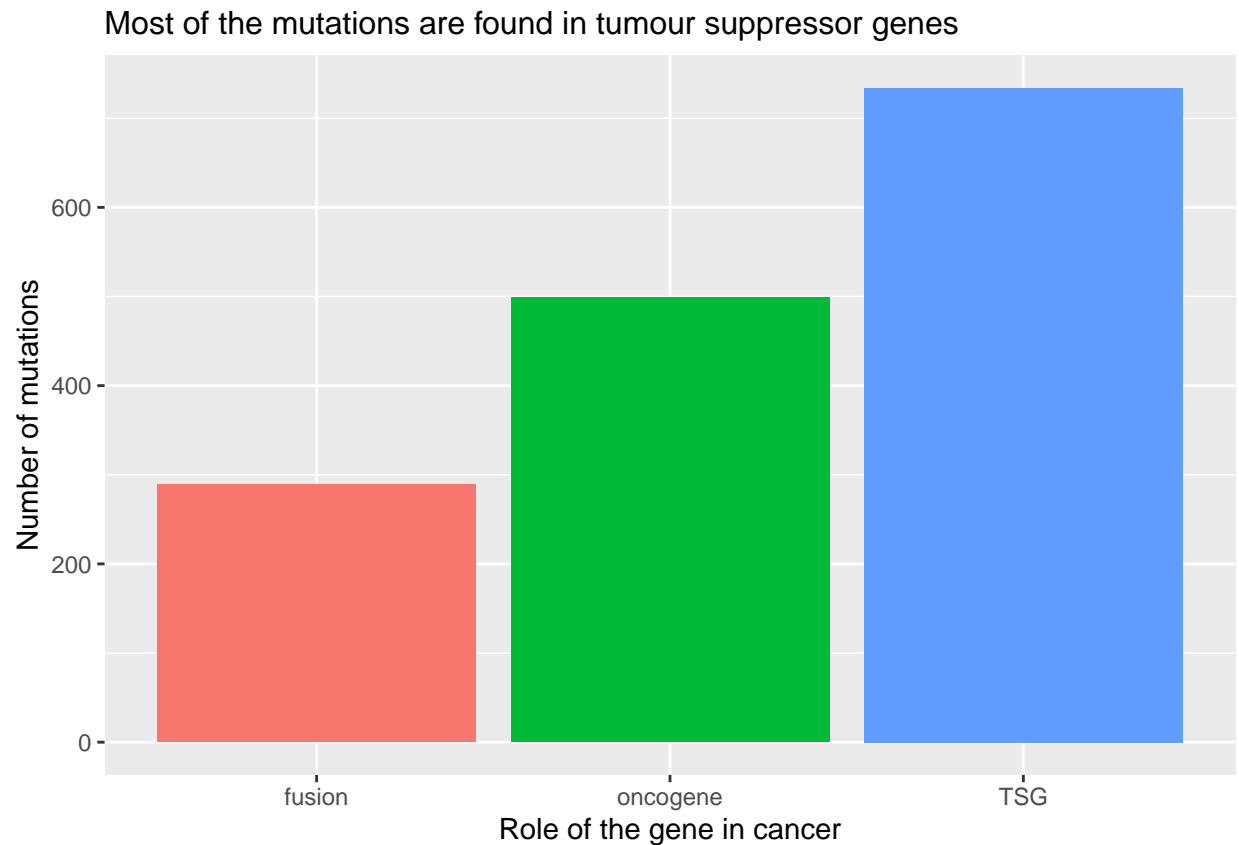
Most of the mutations are found in tumour suppressor genes

Figure 11: Among the mutations found in cancer patients with cryptic STMN2 expression, majority of the mutations are in tumour suppressor genes.

Table 3: Almost half of the cancer genes mutated in patients with enrichment of cryptic STMN2 events are tumour suppressor genes.

role_in_cancer	n	percent
TSG	734	0.4822602
oncogene	499	0.3278581
fusion	289	0.1898817