

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))
```

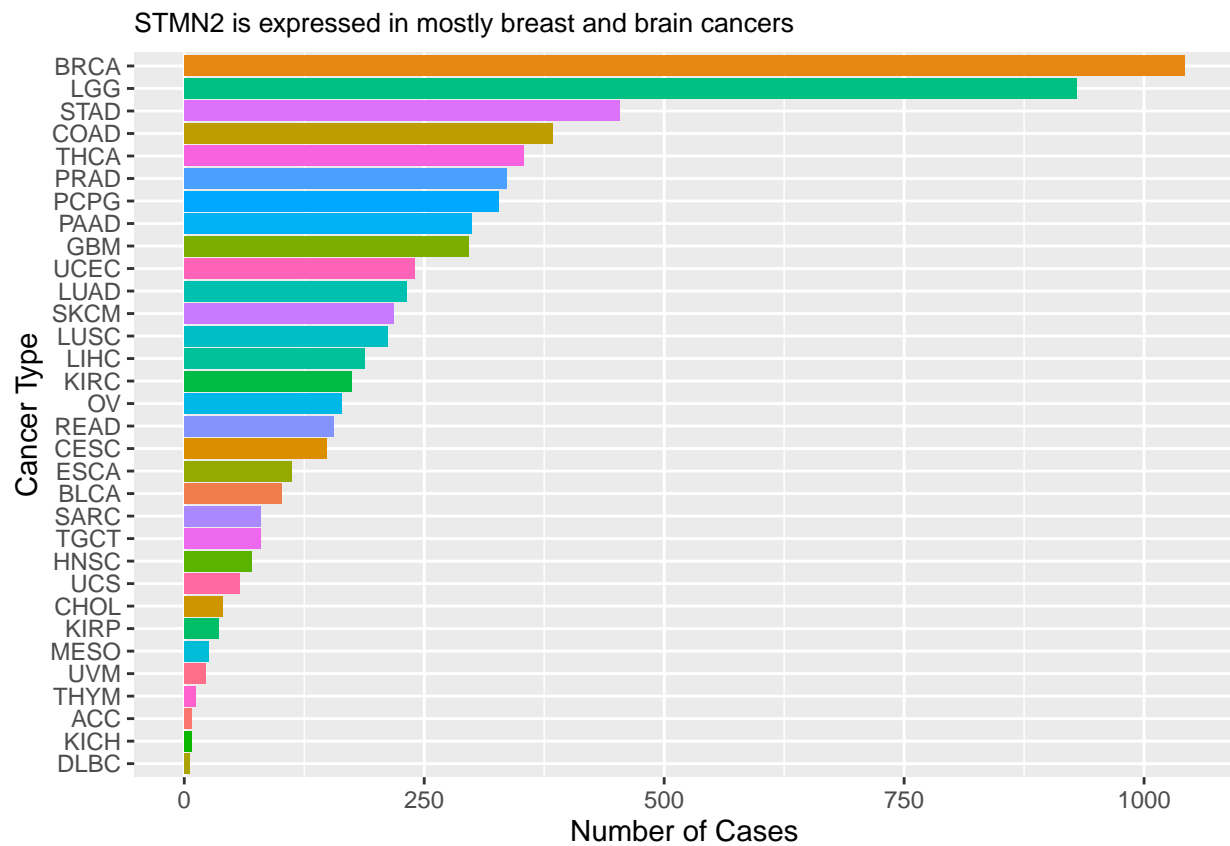


Figure 1: STMN2 events are found mostly in breast and brain cancer patients. BRCA = breast cancer; LGG = low-grade gliomas (brain tumours)

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))
```

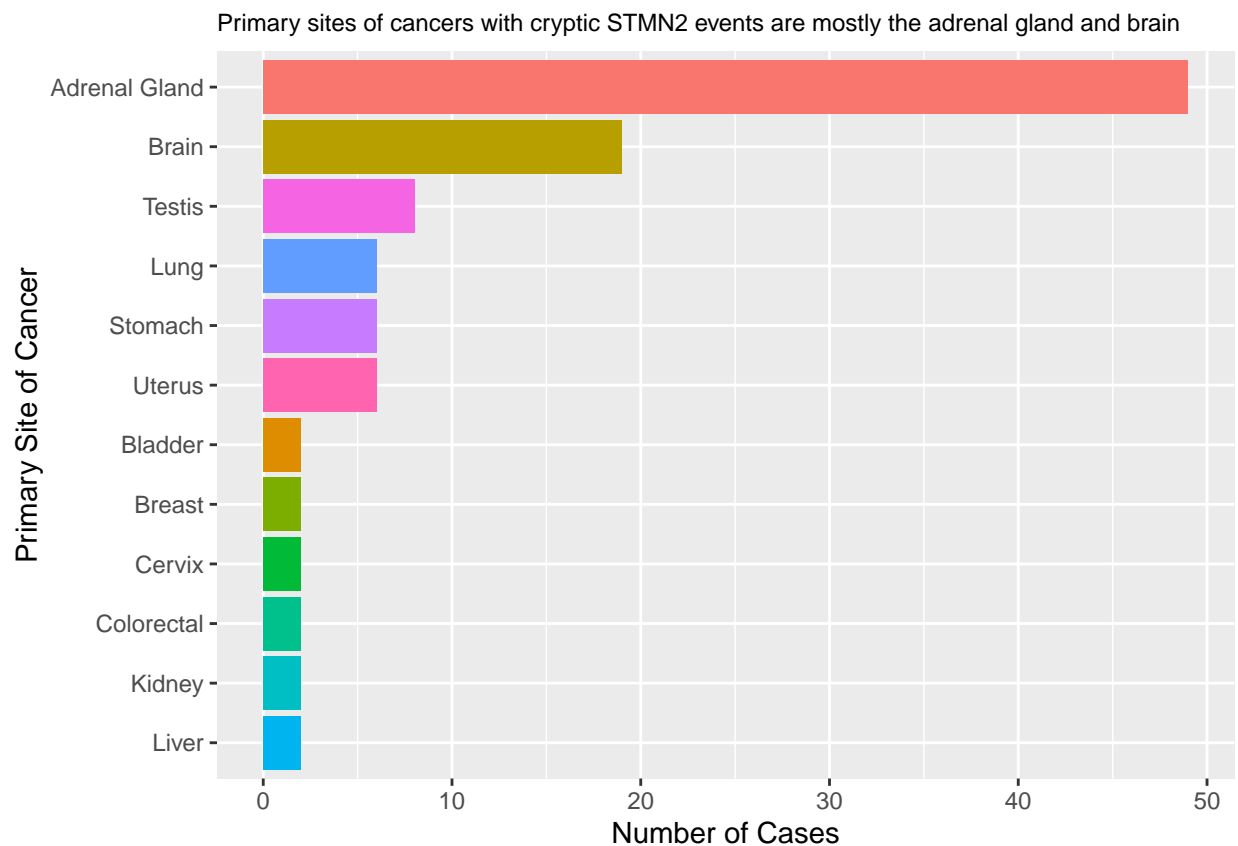


Figure 2: Cryptic STMN2 events are found mostly in the adrenal gland and brain.

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))
```

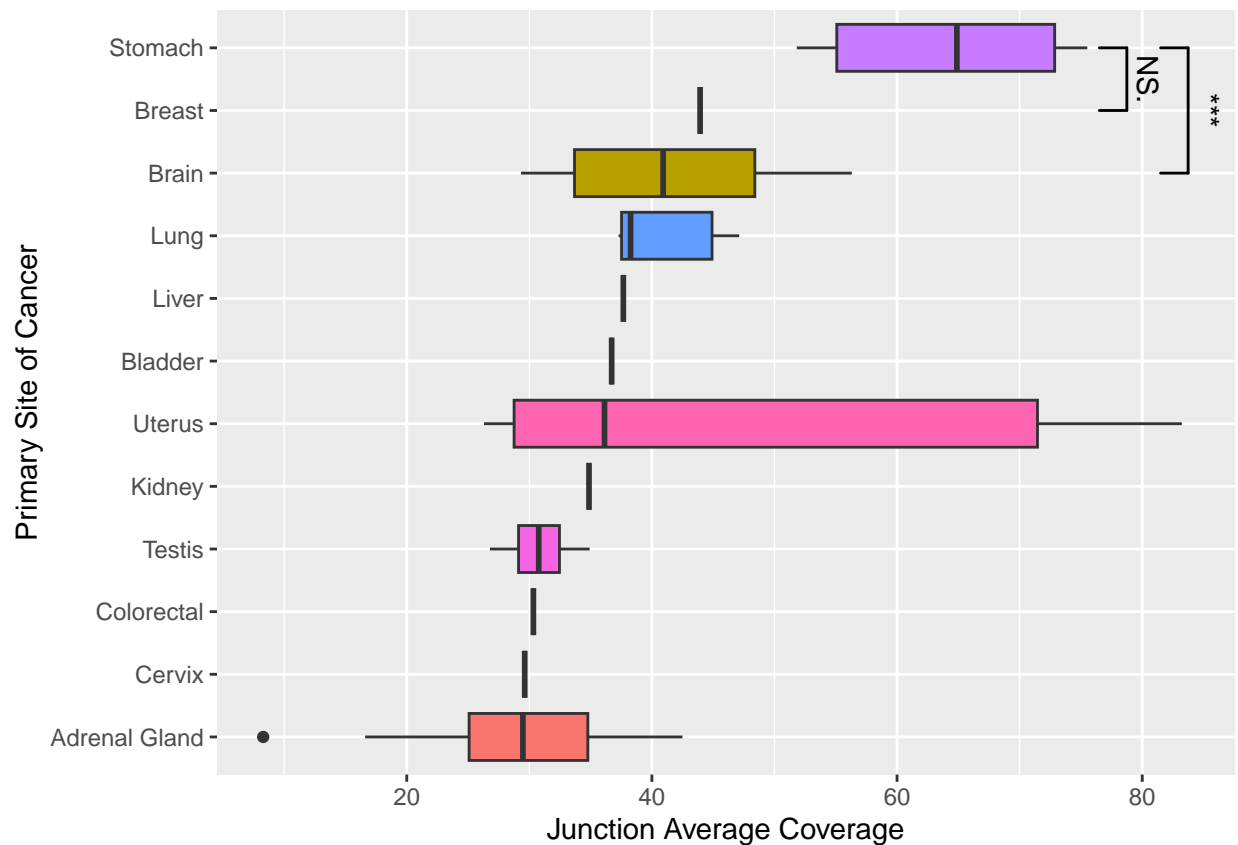


Figure 3: Stomach and breast cancers are the most deeply sequenced.

```
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 = "Junction Average Coverage",
  y = "Primary Site of Cancer",
) +
theme(legend.position = "none", plot.title = element_text(size=10))

```

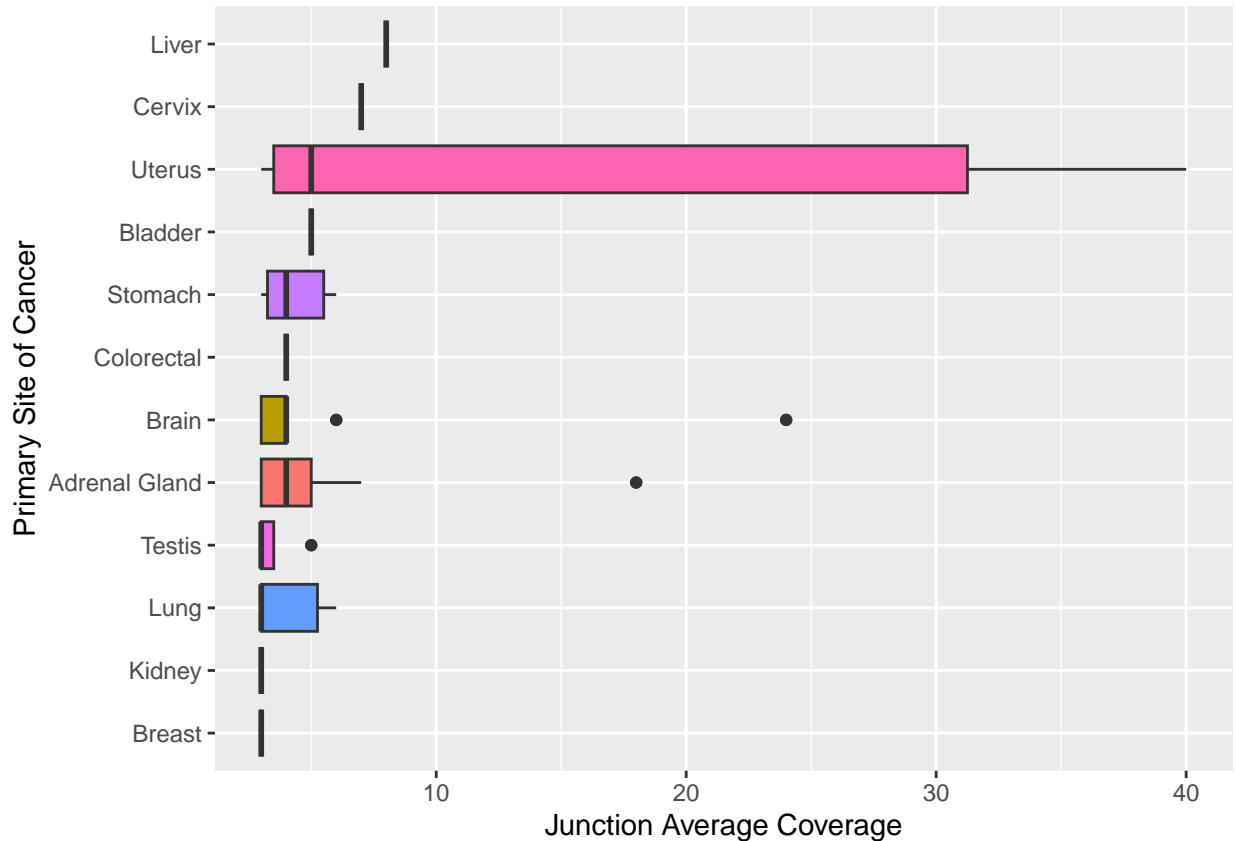


Figure 4:

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 supporting 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

cancer_type	n	percent
Stomach Adenocarcinoma	6	0.0566038
Brain Lower Grade Glioma	4	0.0377358
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

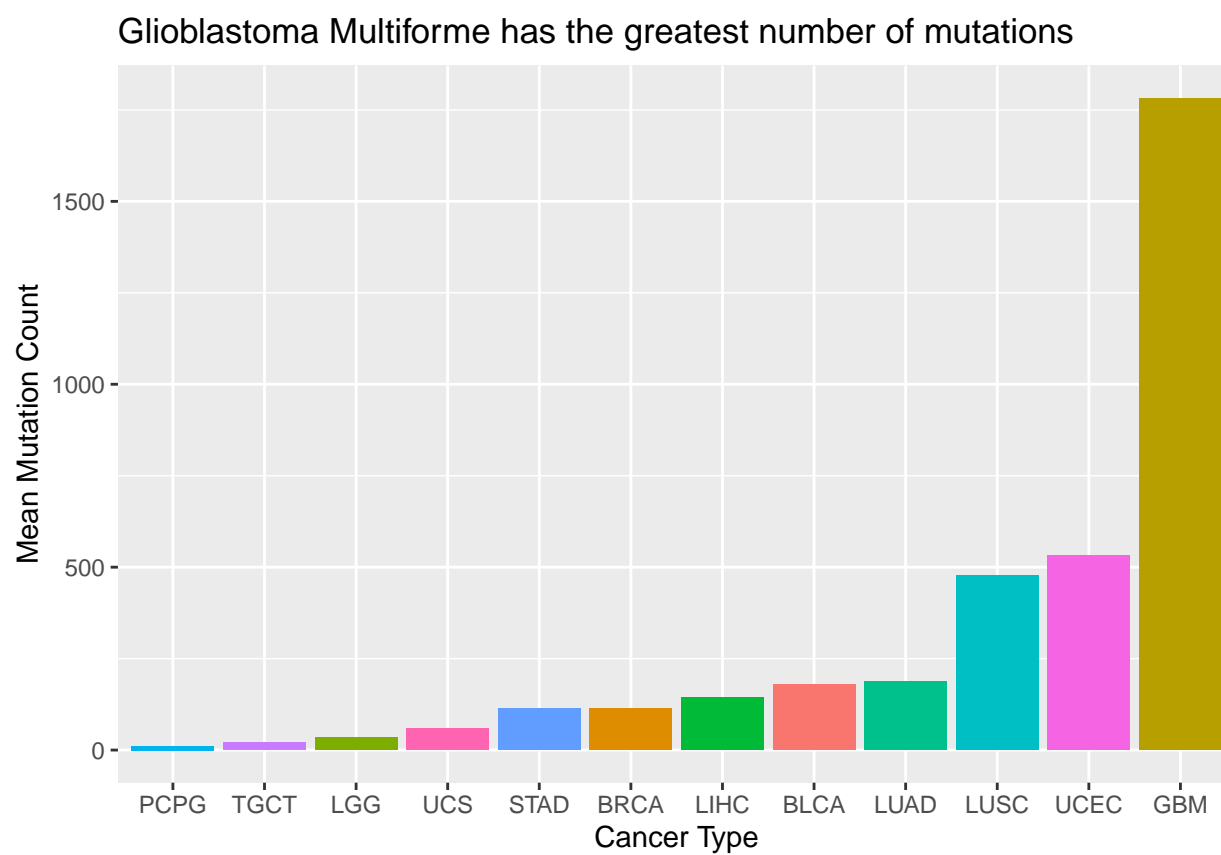


Figure 5: Among cancer patients with STMN2 cryptic expression, GBM (brain cancer) has the greatest number of mutations.

Mutation Data of one patient

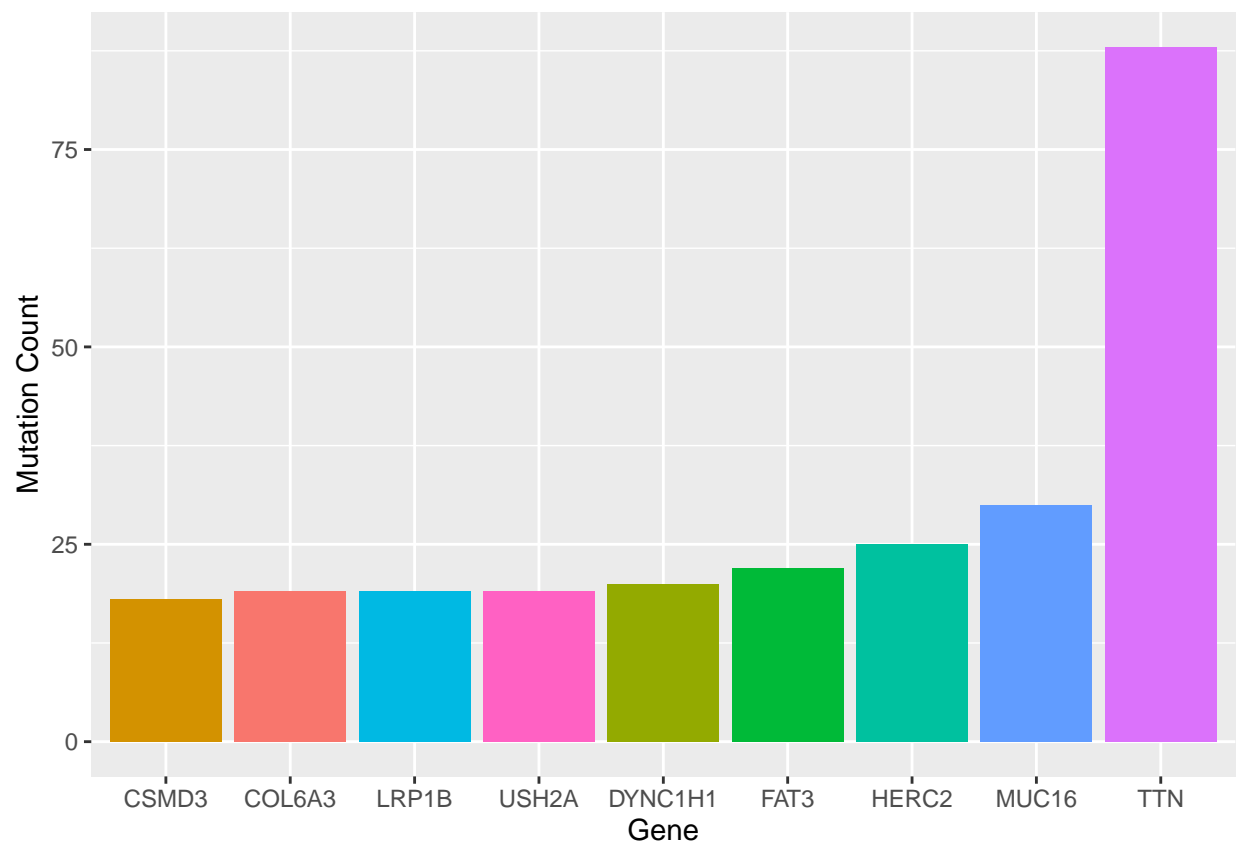


Figure 6:

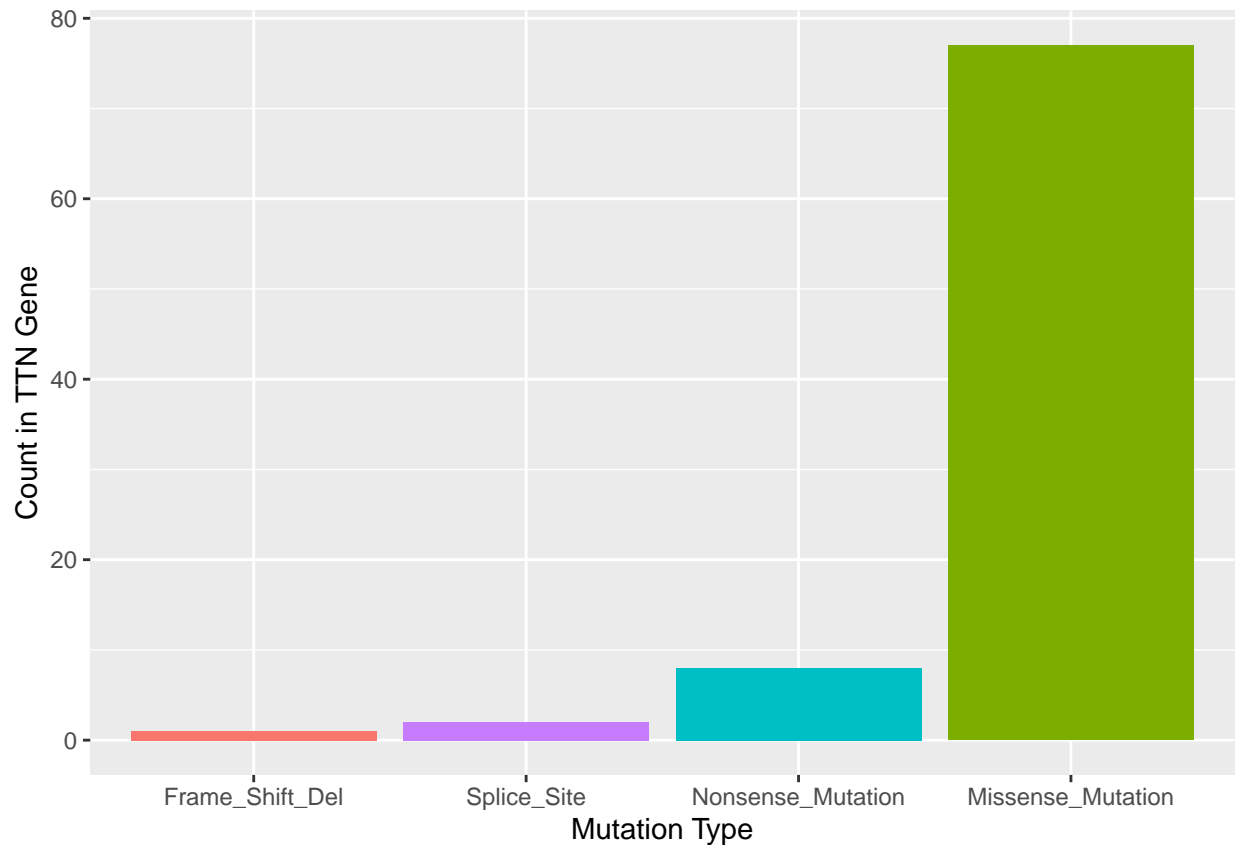


Figure 7:

STMN2 cryptic expression

```
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() |>
  ggplot(aes(x = stmn2_cryptic_coverage, y = mutation_count)) +
  labs(
    x = "Number of STMN2 cryptic events",
    y = "Mutation Count"
  ) +
  geom_point()
```

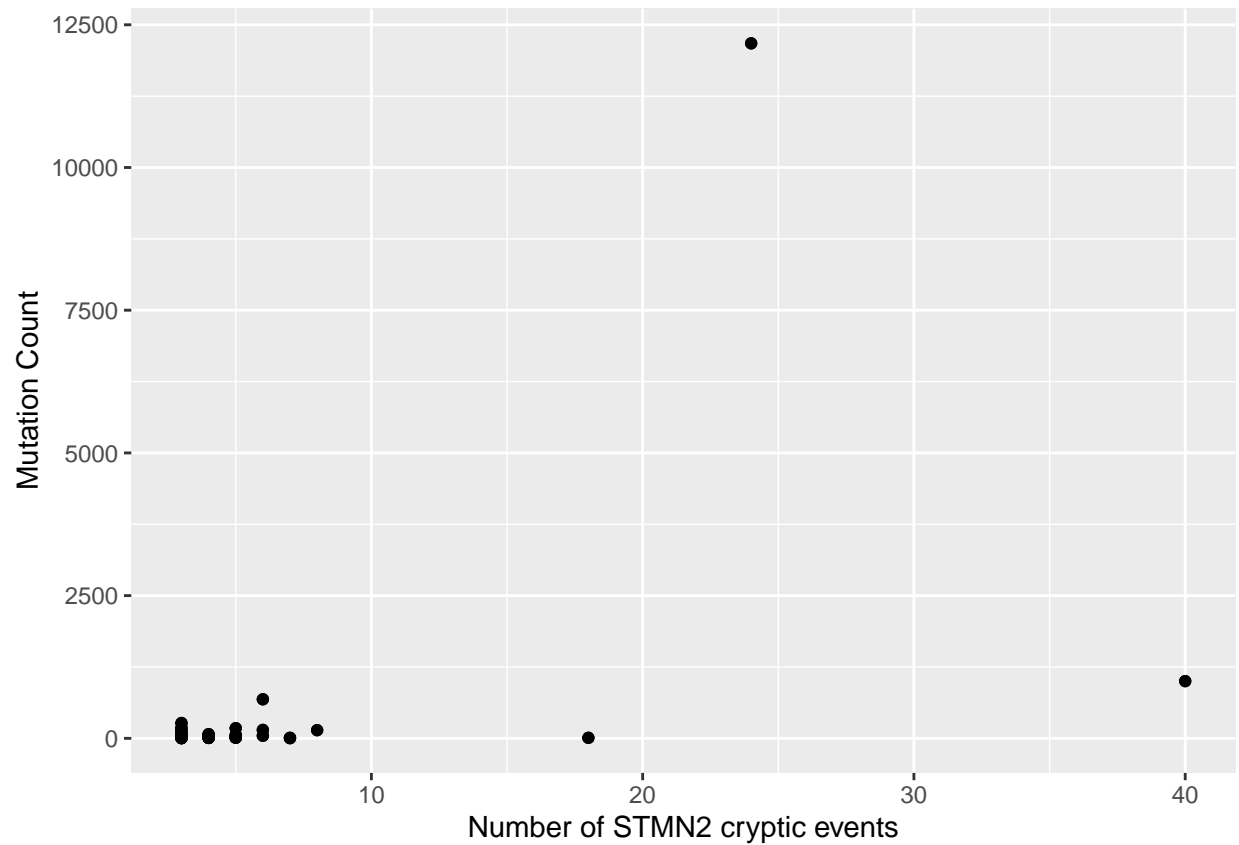



Figure 8:

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")
```

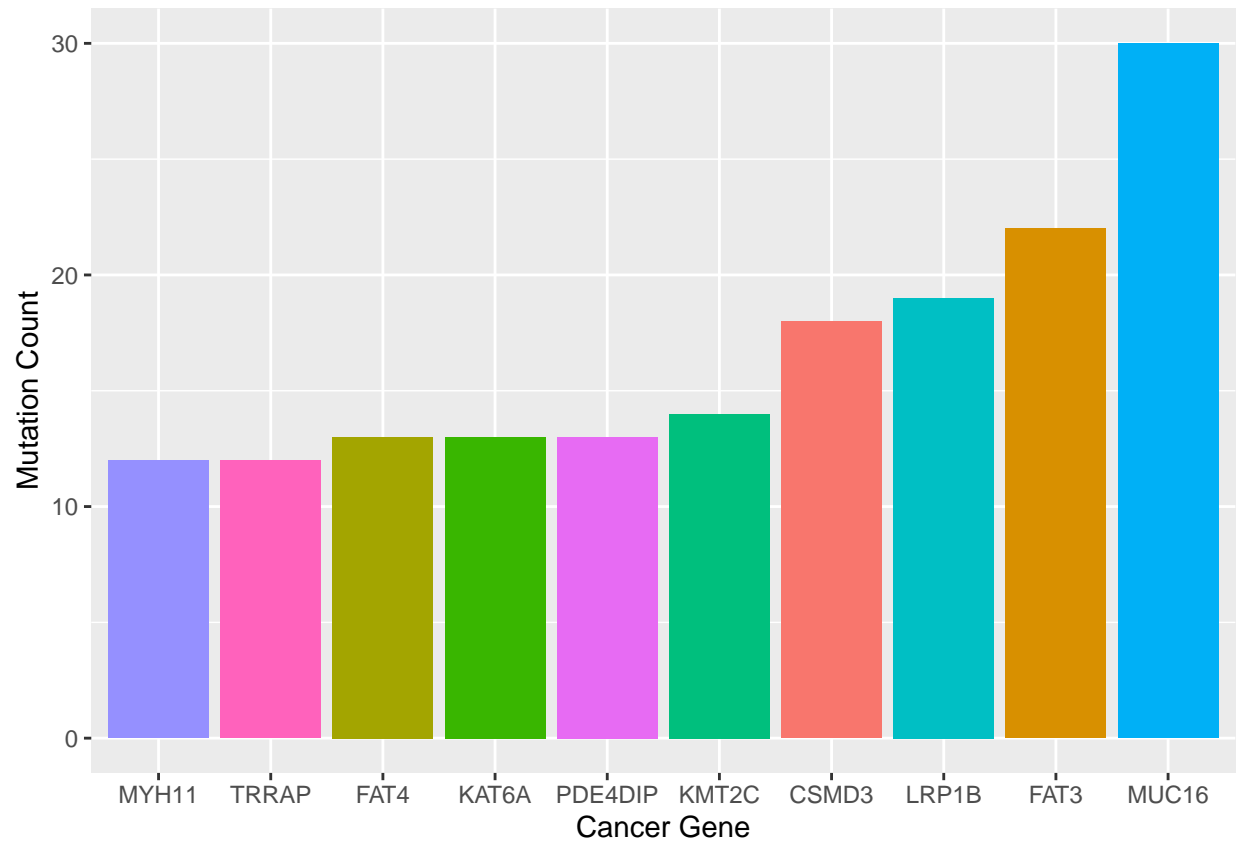


Figure 9: Ovarian cancer gene MUC16 has the greatest number of mutations.

```
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")
```

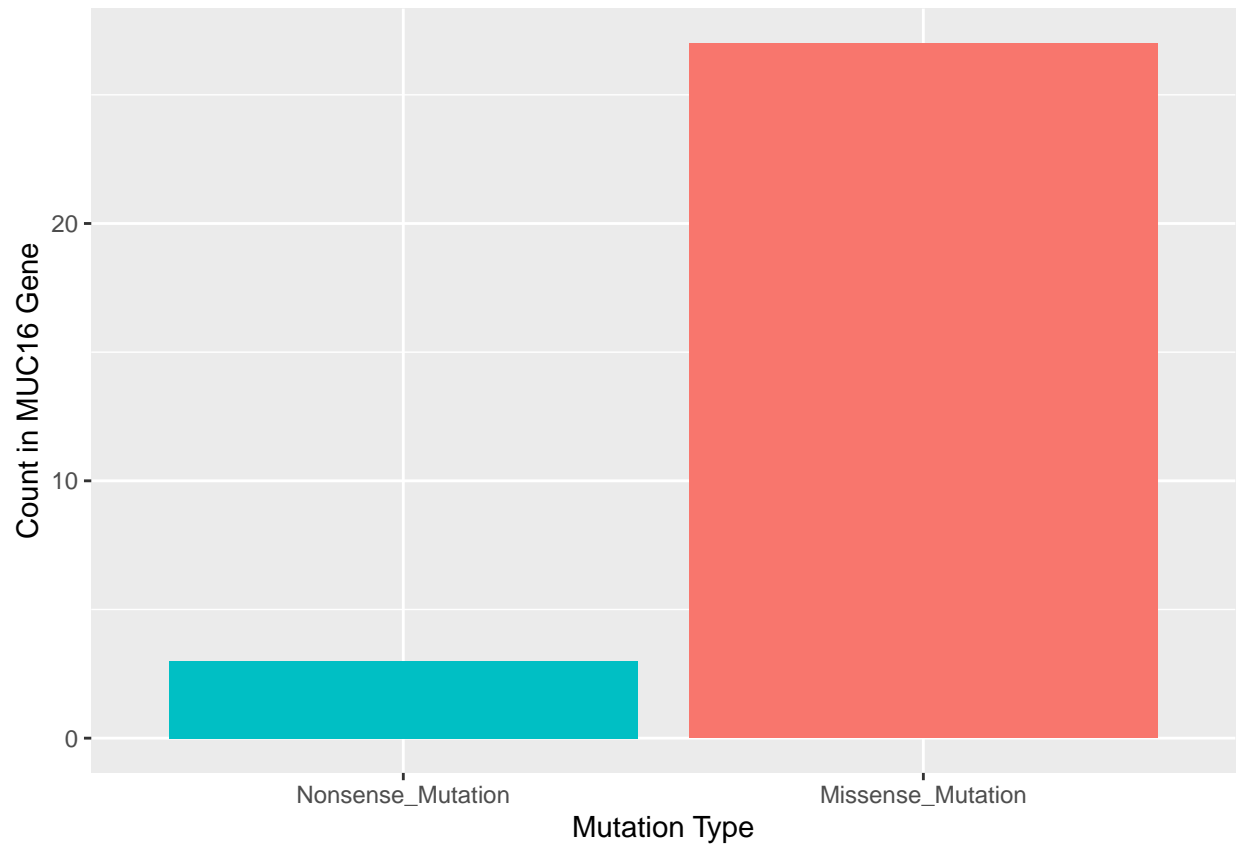


Figure 10: Ovarian cancer gene MUC16 has mostly missense mutations.

Types of cancer genes

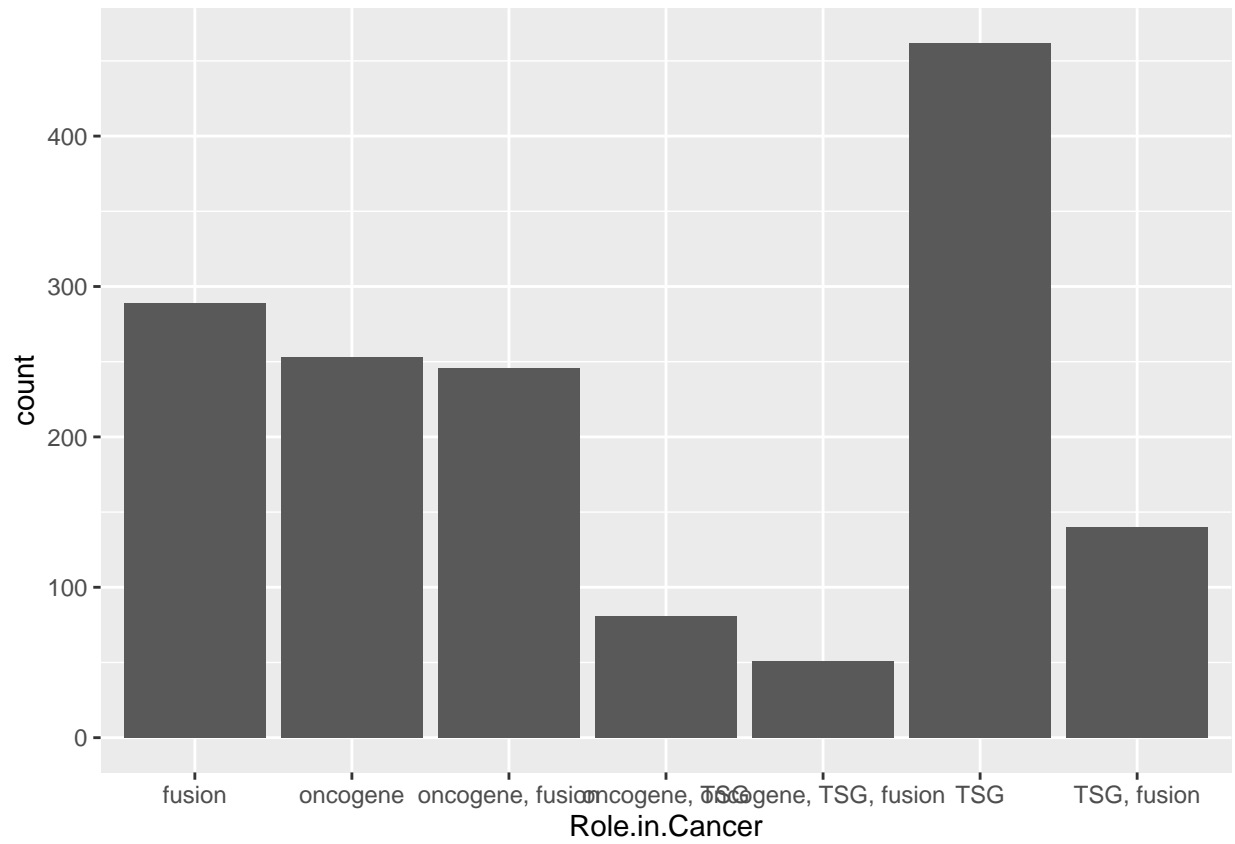


Table 3: 45% of cancer genes with cryptic STMN2 expression are tumour suppressor genes.

TSG	n	percent
no	901	0.5510703
yes	734	0.4489297

Table 4: 39% of cancer genes with cryptic STMN2 expression are oncogenes.

oncogene	n	percent
no	1004	0.6140673
yes	631	0.3859327

Table 5: 44% of cancer genes with cryptic STMN2 expression are gene fusions.

fusion	n	percent
no	909	0.5559633

fusion	n	percent
yes	726	0.4440367

TCGA biolinks