

Enhancer Count Analysis

Avery Holloman

2024-08-13

```
# Libraries
library(ggplot2)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##   filter, lag
```

```
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

```

# Fictitious dataset of the abilities of enhancer counts from different sources
data <- data.frame(
  enhancer_length = rep(seq(1, 5, 0.1), 7),
  enhancer_count = c(
    dnorm(seq(1, 5, 0.1), mean = 2.5, sd = 0.5) * 10^4, # Ensembl
    dnorm(seq(1, 5, 0.1), mean = 2.8, sd = 0.6) * 10^4, # FANTOM
    dnorm(seq(1, 5, 0.1), mean = 2.6, sd = 0.4) * 10^3, # VISTA
    dnorm(seq(1, 5, 0.1), mean = 3.2, sd = 0.5) * 10^3, # ENCODE
    dnorm(seq(1, 5, 0.1), mean = 2.9, sd = 0.5) * 10^3, # All sources
    dnorm(seq(1, 5, 0.1), mean = 2.5, sd = 0.6) * 10^3, # All clusters
    dnorm(seq(1, 5, 0.1), mean = 3.0, sd = 0.5) * 10^4 # Custom data
  ),
  source = factor(rep(c("Ensembl", "FANTOM", "VISTA", "ENCODE", "All sources", "All clusters",
"Custom"), each = length(seq(1, 5, 0.1))))
)

# I utilized ggplot2 to show the data
ggplot(data, aes(x = enhancer_length, y = enhancer_count, color = source)) +
  geom_line(size = 1) +
  scale_y_log10() +
  scale_x_log10() +
  labs(x = "Enhancer length (bp), log10", y = "Enhancer count, log10", color = "Source") +
  theme_minimal() +
  theme(
    text = element_text(size = 12),
    legend.position = "top",
    legend.title = element_blank(),
    plot.title = element_text(hjust = 0.5)
  ) +
  ggtitle("Enhancer Count by Length across Sources")

```

```

## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.

```

Enhancer Count by Length across Sources

