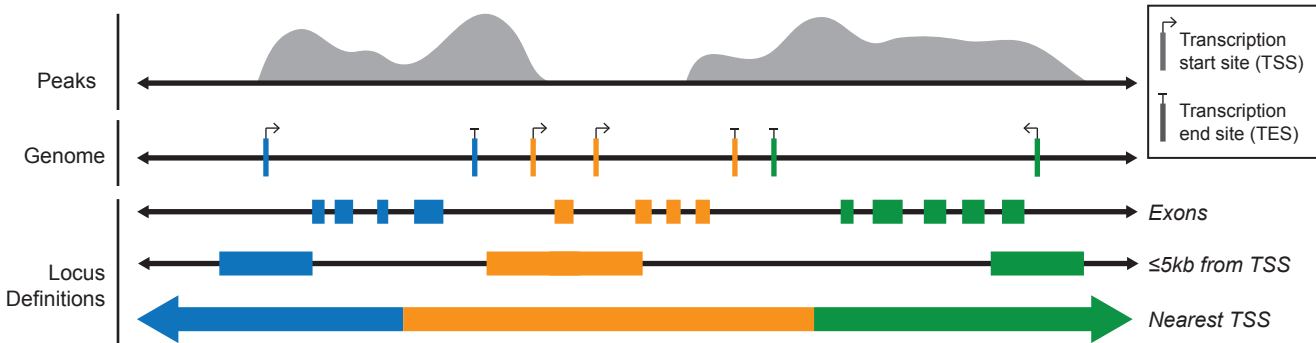


## 1. Assign peaks to genes



## 2. Determine proportion of peak overlap with locus

Gene	Locus length	Proportion covered ( $r$ )
APP	433,986	0.2087
CTCF	86,254	0.4201
...	...	...

## 3. Test for gene set enrichment

$$\text{geneset} = b_0 + b_1 r + \text{SS}(\log_{10} L)$$

Logistic regression model

- Adjust for locus length with smoothing spline (SS)
- Estimate effect size ( $b_1$ ) of locus coverage ( $r$ )
- Test  $H_0: b_1 = 0$  versus  $H_1: b_1 \neq 0$

## 4. Summarize data and enrichment results

Geneset ID	p-value	FDR	Odds Ratio
GO:0019899	1.01E-20	8.80E-18	3.351
GO:0031981	3.07E-18	1.51E-15	2.354
GO:0005654	9.57E-16	2.36E-13	2.469
...	...	...	...