

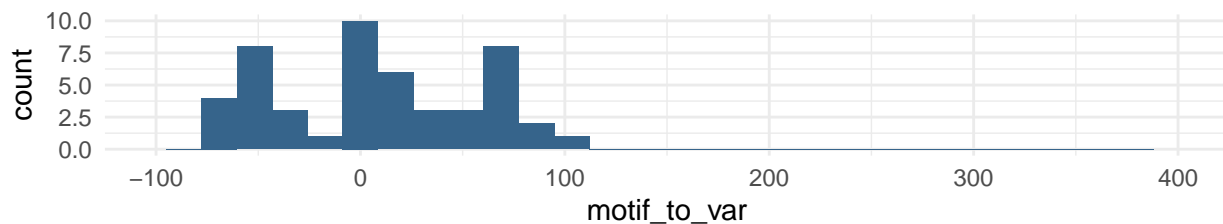
Location of MPRA variants in repeats

Grace Hansen

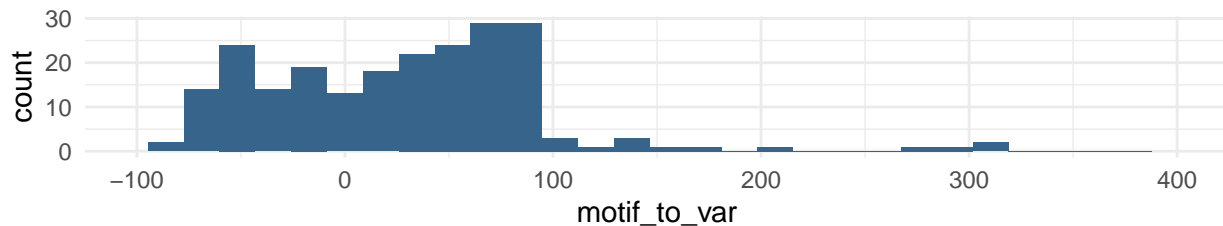
06/29/2019

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## Warning: Removed 10 rows containing non-finite values (stat_bin).  
## Warning: Removed 2 rows containing missing values (geom_bar).  
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## Warning: Removed 144 rows containing non-finite values (stat_bin).  
## Warning: Removed 2 rows containing missing values (geom_bar).  
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## Warning: Removed 913 rows containing non-finite values (stat_bin).  
## Warning: Removed 2 rows containing missing values (geom_bar).
```

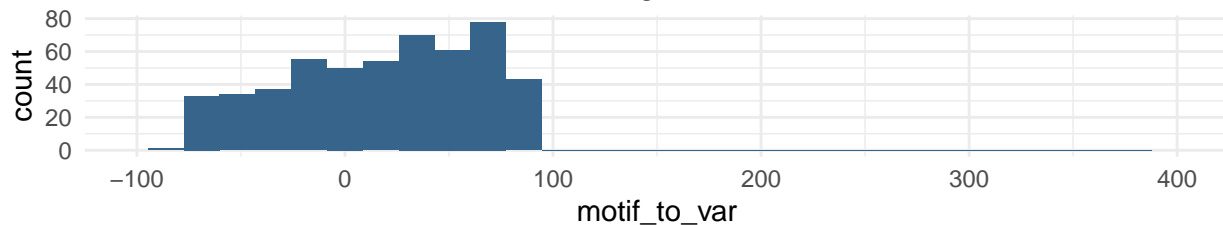
Distance from motif to variant, EMVars



Distance from motif to variant, enhancers



Distance from motif to variant, nonsignificant variants



Overlap between TF motifs and Alus in EMVars

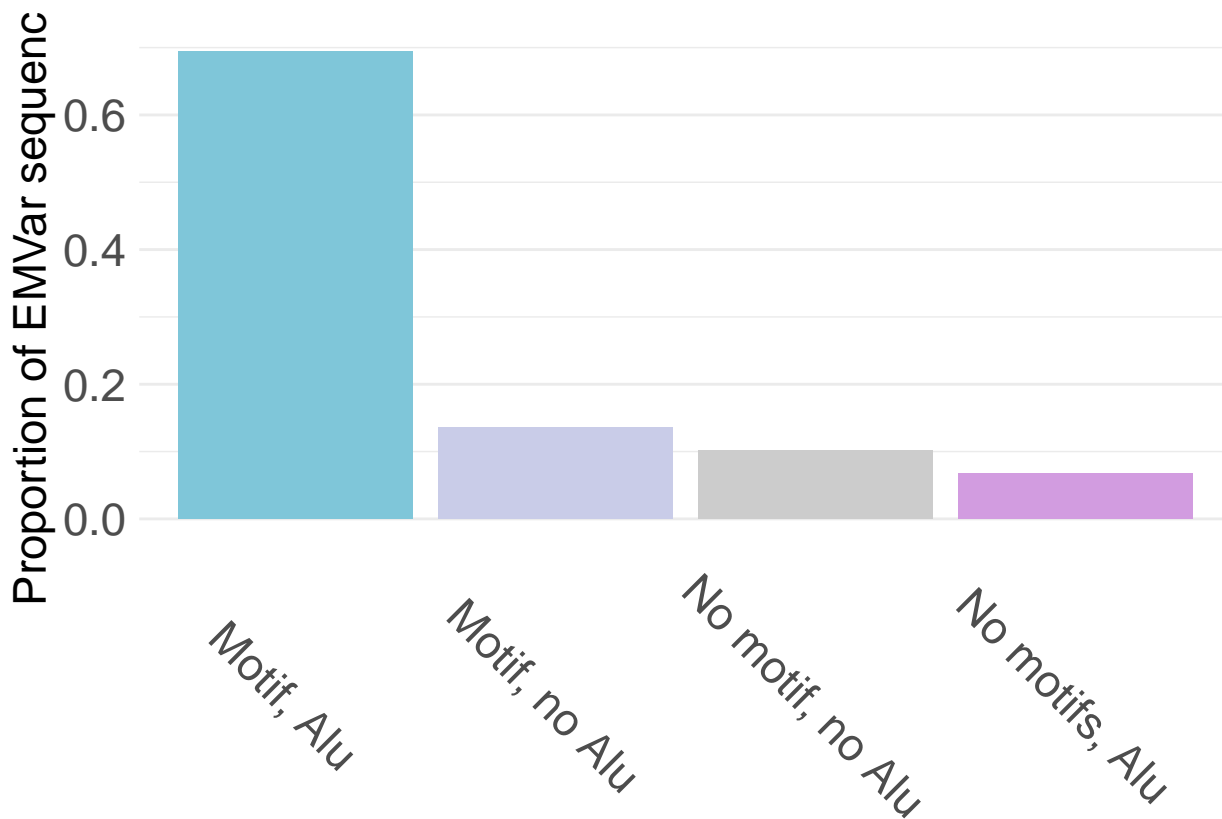
Are EMVar sequences containing Alus more likely to contain a shared TF motif?

Table 1: Observed prevalence of motifs and Alus in EMVars

	Motif	No motif
Alu	41	4
no Alu	8	6

Table 2: Expected prevalence of motifs and Alus in EMVars

	Motif	No motif
Alu	37.37288	7.627119
no Alu	11.62712	2.372881



Overlap between TF motifs and Alus in enhancers

Are enhancer sequences containing Alus more likely to contain a shared TF motif?

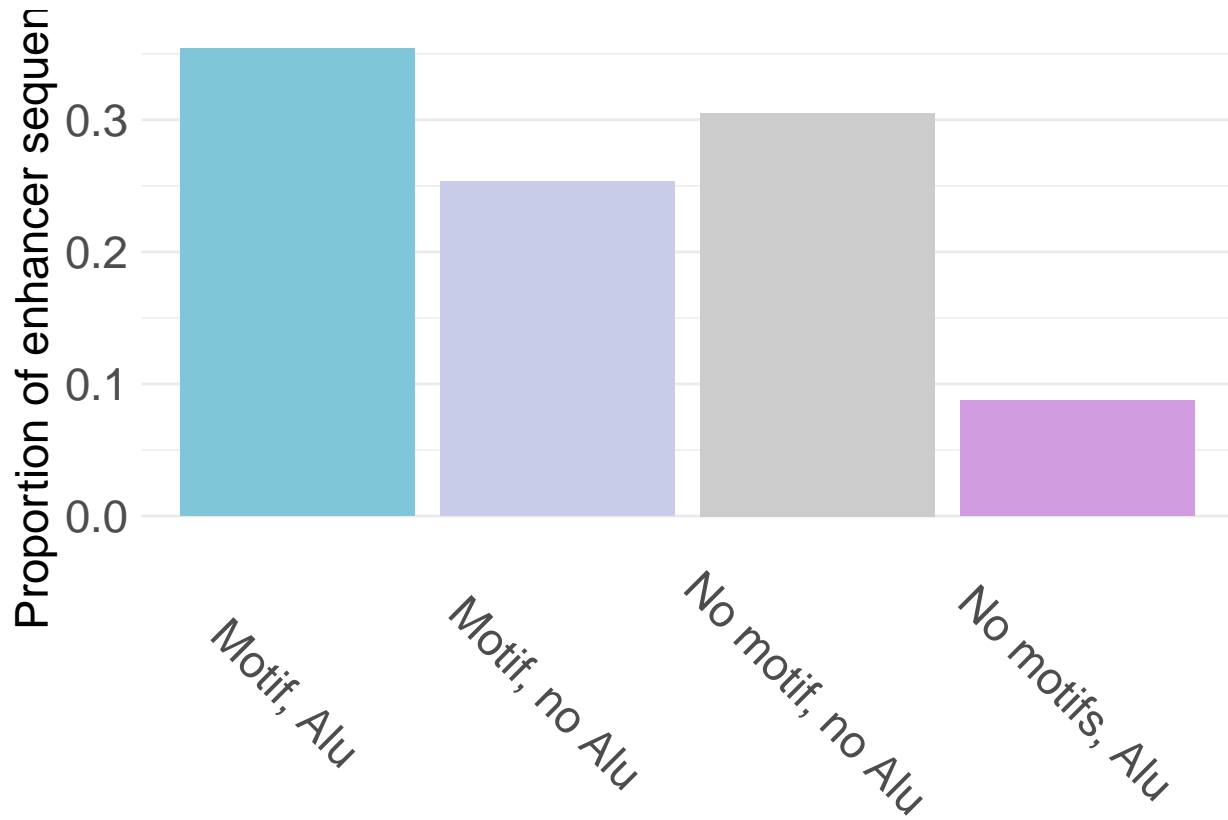
Table 3: Observed prevalence of motifs and Alus in enhancers

	Motif	No motif
Alu	130	32

	Motif	No motif
no Alu	93	112

Table 4: Expected prevalence of motifs and Alus in enhancers

	Motif	No motif
Alu	98.43597	63.56403
no Alu	124.56403	80.43597



Overlap between TF motifs and Alus in all significant sequences

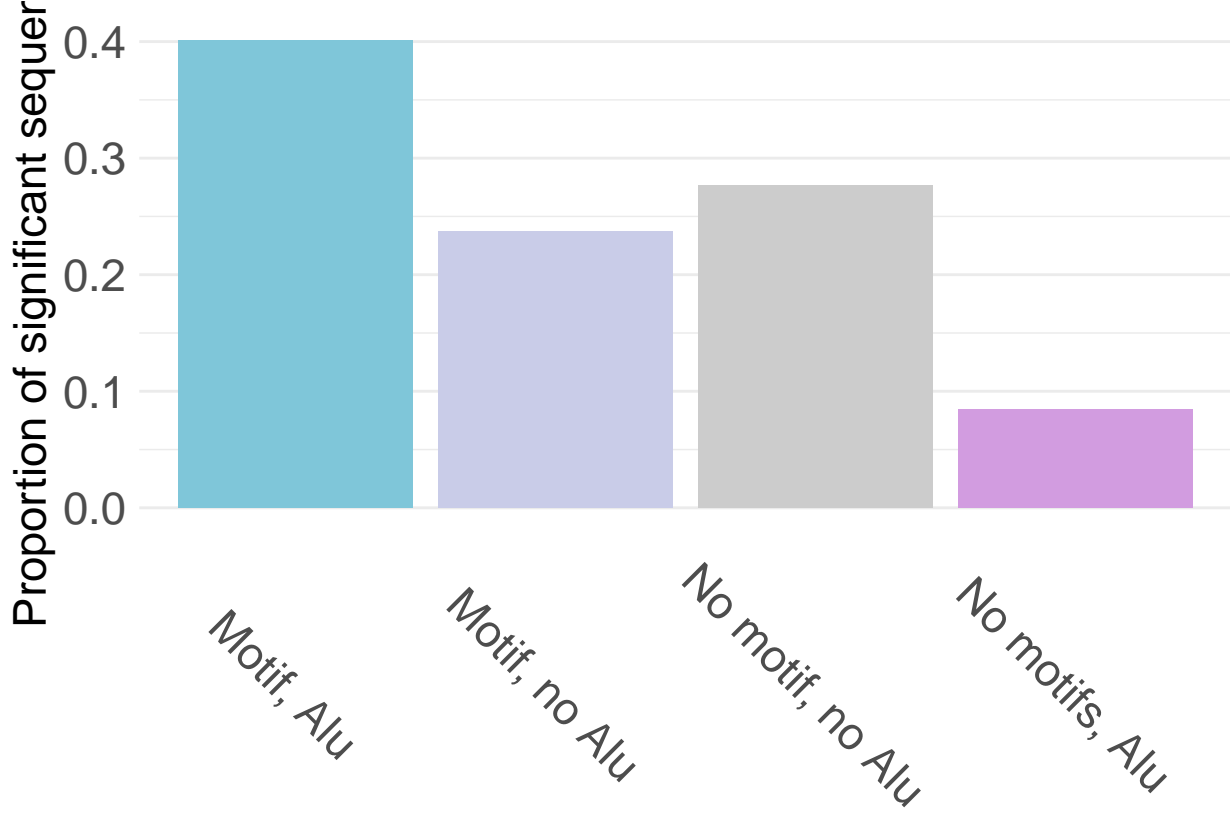
Are enhancer sequences containing Alus more likely to contain a shared TF motif?

Table 5: Observed prevalence of motifs and Alus in significant sequences

	Motif	No motif
Alu	171	36
no Alu	101	118

Table 6: Expected prevalence of motifs and Alus in significant sequences

	Motif	No motif
Alu	132.169	74.83099
no Alu	139.831	79.16901



Overlap between TF motifs and Alus in nonsignificant sequences

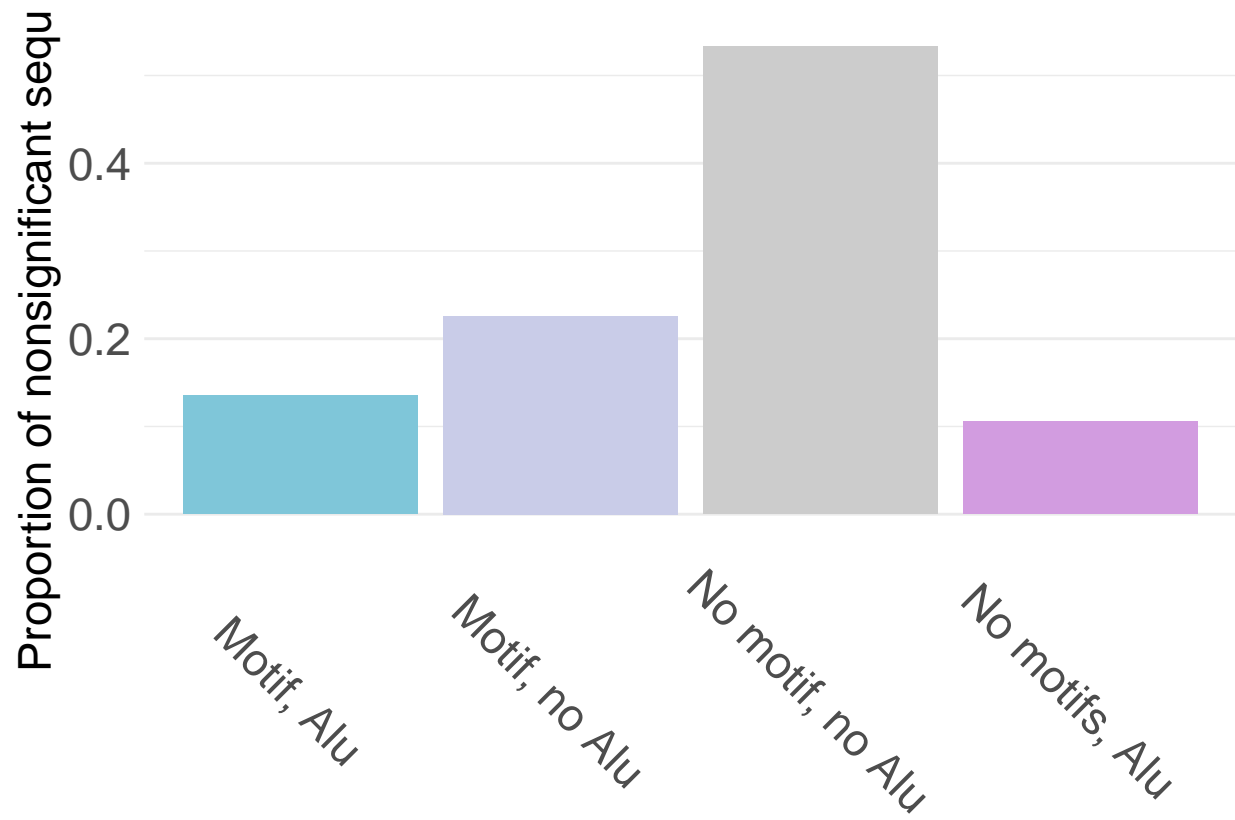
Are enhancer sequences containing Alus more likely to contain a shared TF motif?

Table 7: Observed prevalence of motifs and Alus in nonsig sequences

	Motif	No motif
Alu	193	151
no Alu	323	762

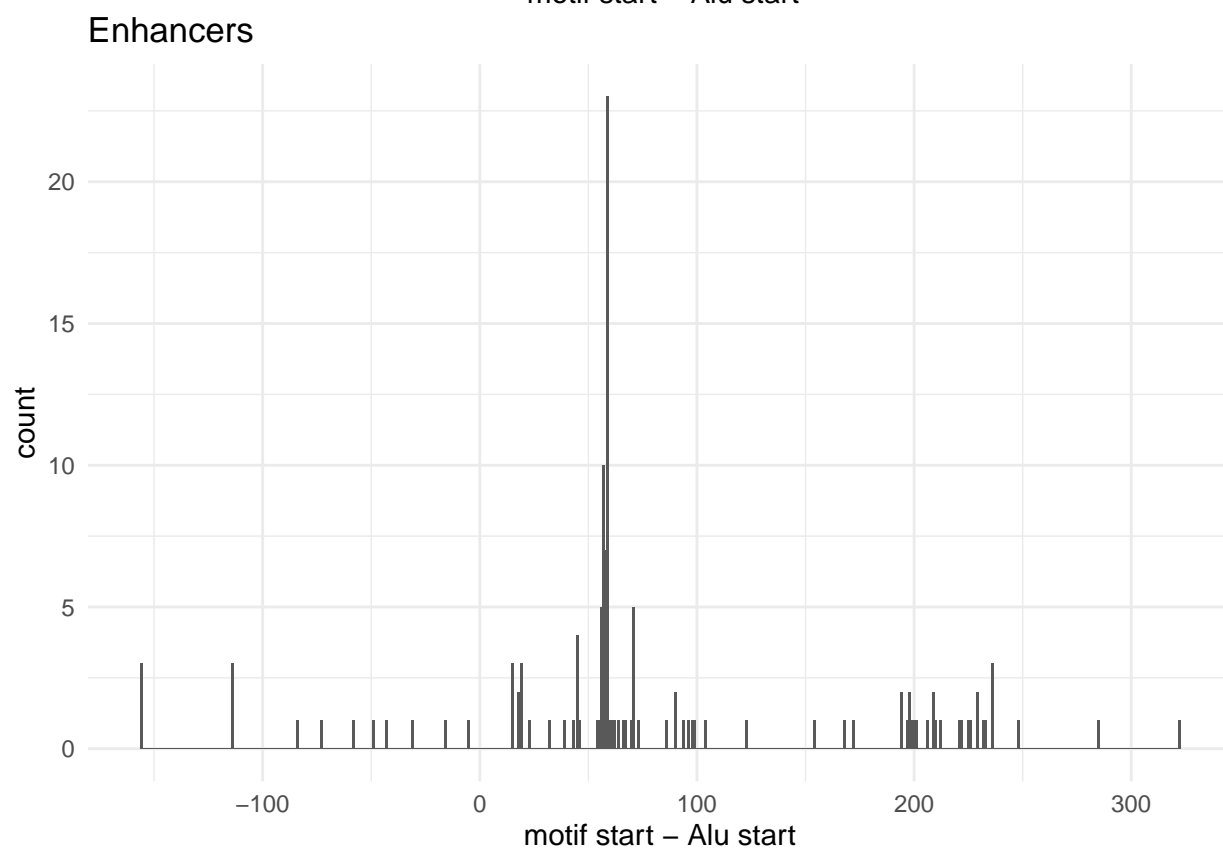
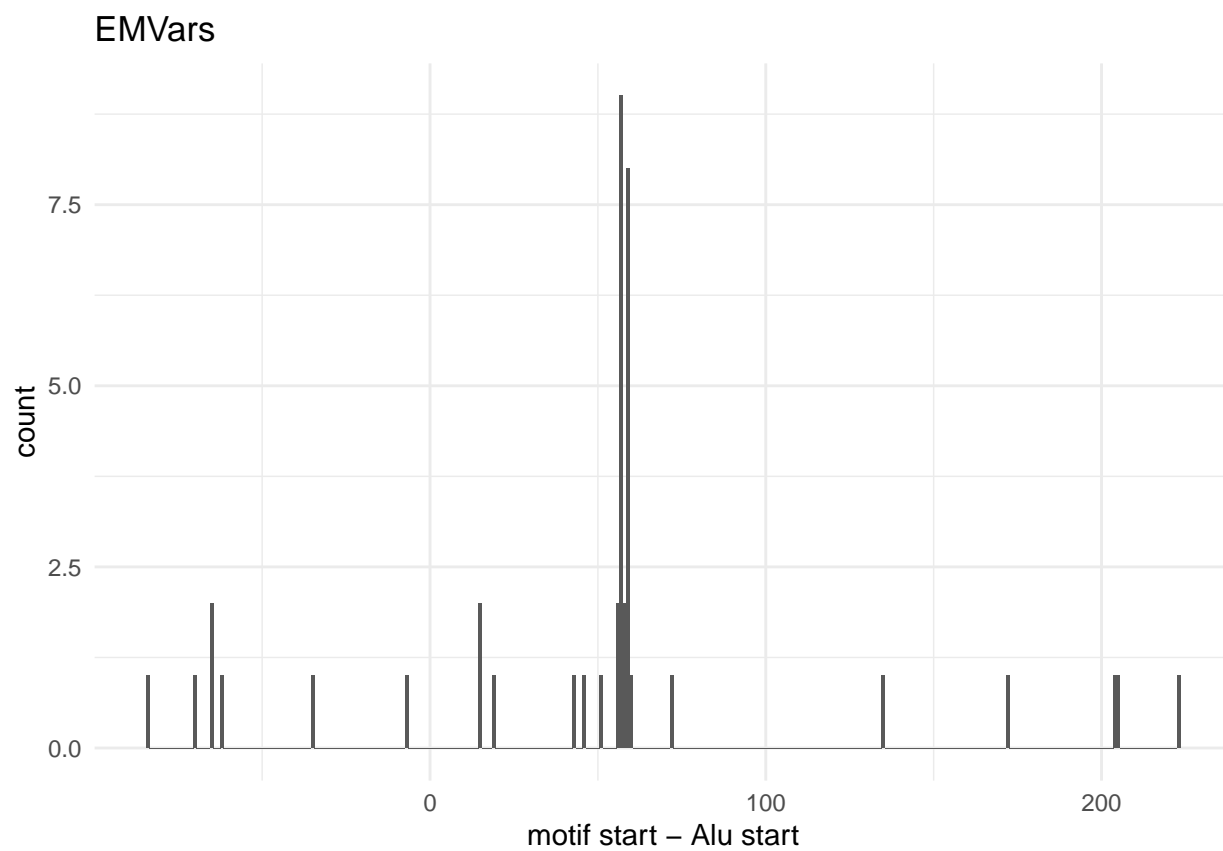
Table 8: Expected prevalence of motifs and Alus in nonsig sequences

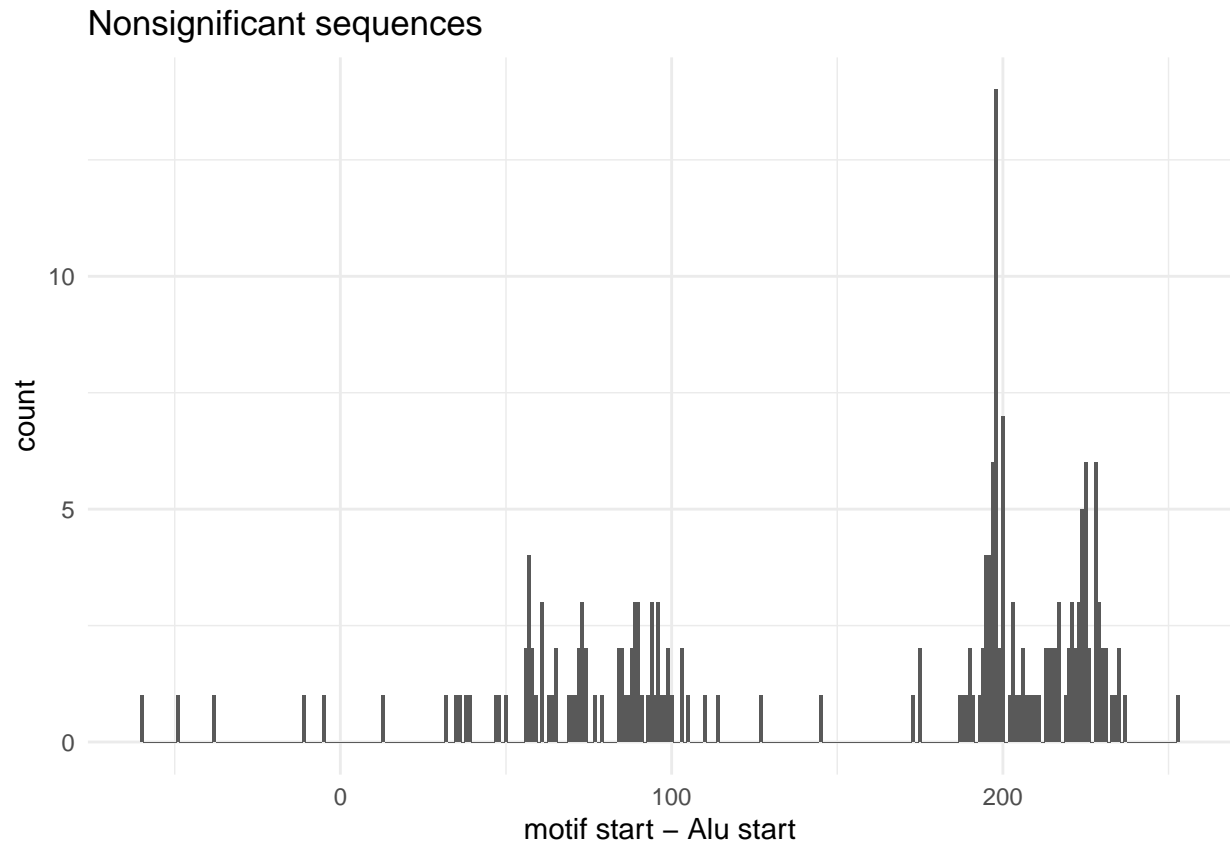
	Motif	No motif
Alu	124.2155	219.7845
no Alu	391.7845	693.2155



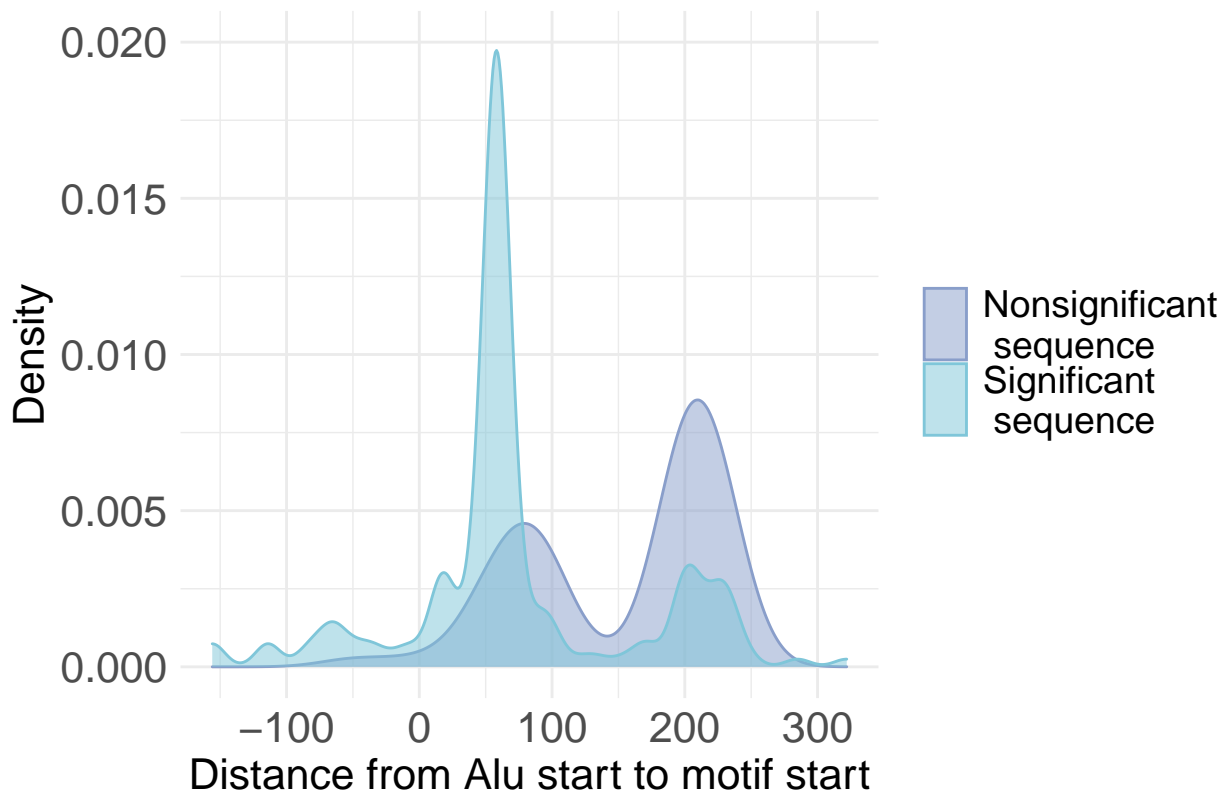
Distance from Alu start to motif start

Are the motifs always a certain distance from the start of a Alu? If so, this might indicated a shared part of the Alu at work.





It looks like significant sequences are likely to contain a motif ~57 bp from the start of a Alu. Let's also visualize this as density plots.

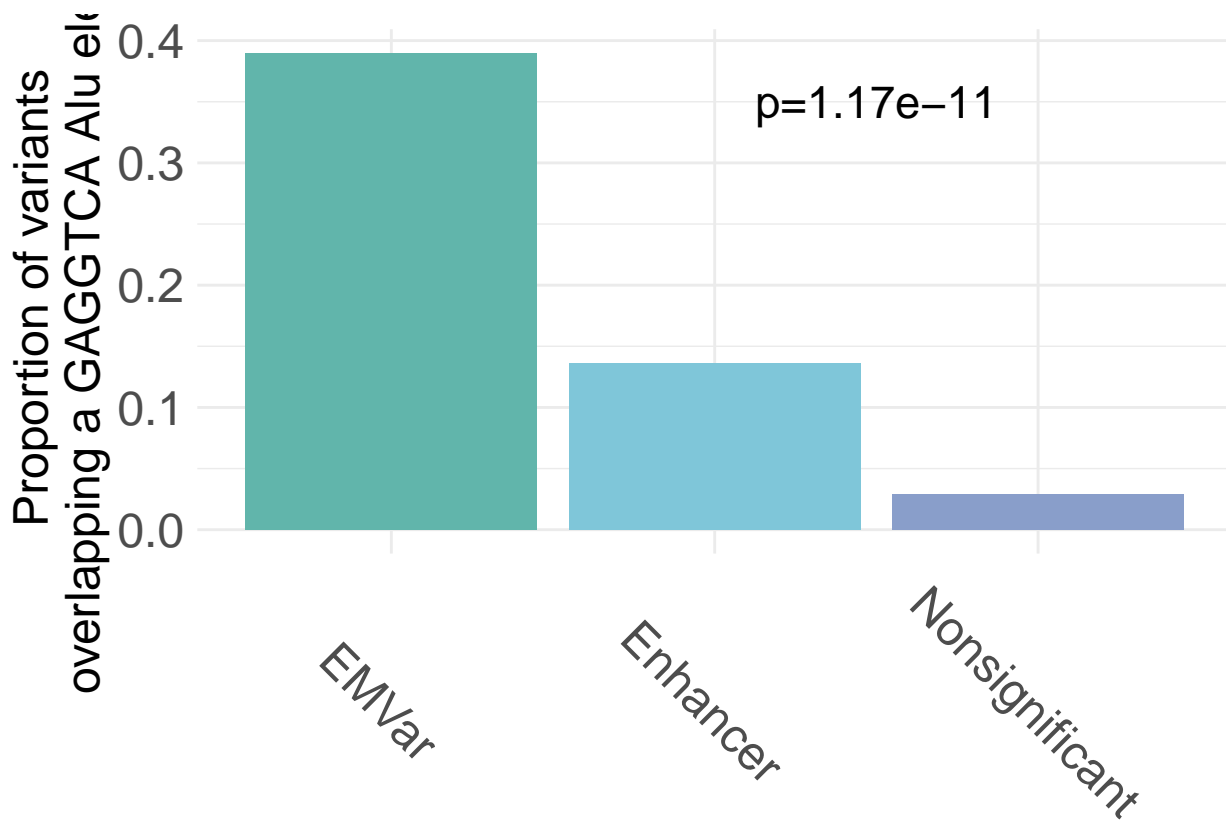


Accessibility of active Alu elements

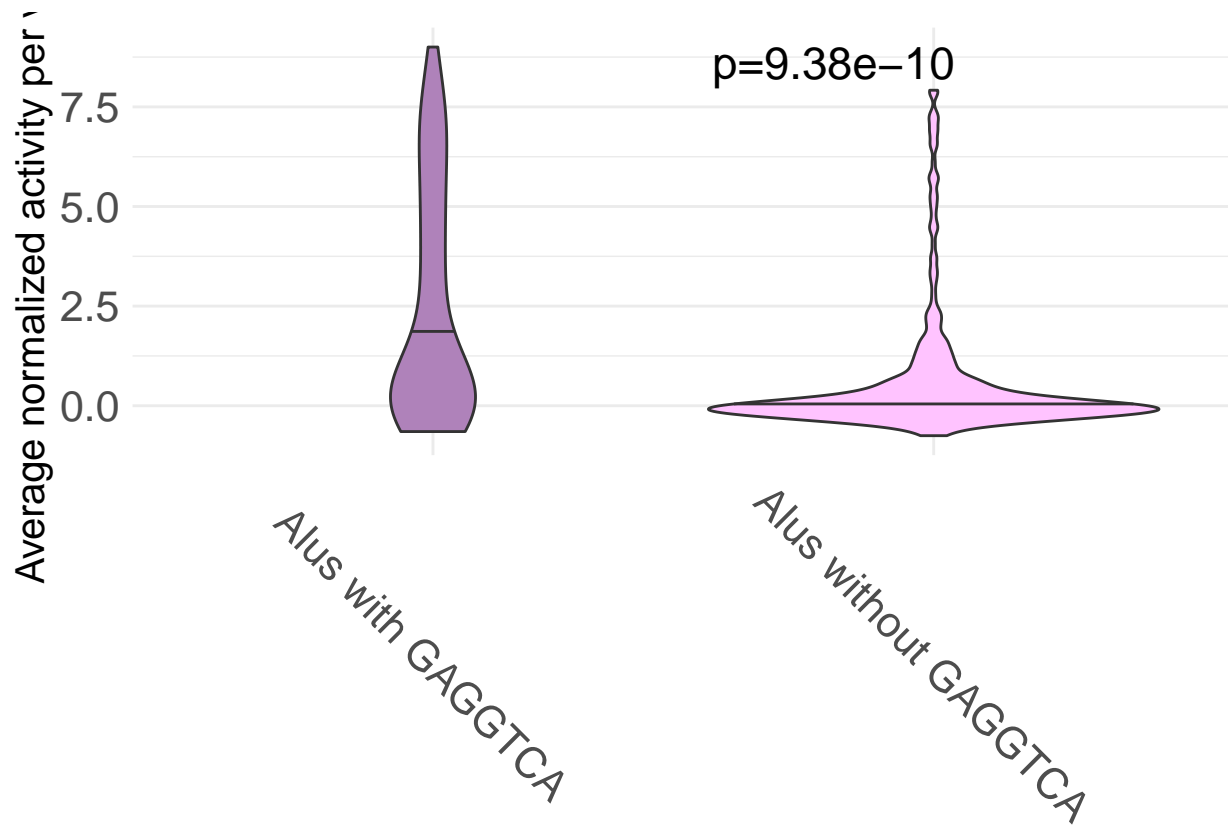
Are the Alu elements that contain an adipogenic motif and drive expression in open chromatin?

```
## [1] "/home/grace/midway/MPRA_repeats"
## [1] "Active repeats overlapping SGBS preadipocyte peaks:"
## character(0)
## [1] "Active repeats overlapping SGBS day2 peaks:"
## character(0)
## [1] "Active repeats overlapping SGBS day8 peaks:"
## character(0)
## [1] "Active repeats overlapping SGBS day16 peaks:"
## character(0)
```

	Alus with GAGGTCA	Alus without GAGGTCA
EMVar	23	22
enhancer	50	112
nonsig	42	302



```
## pdf
## 2
```

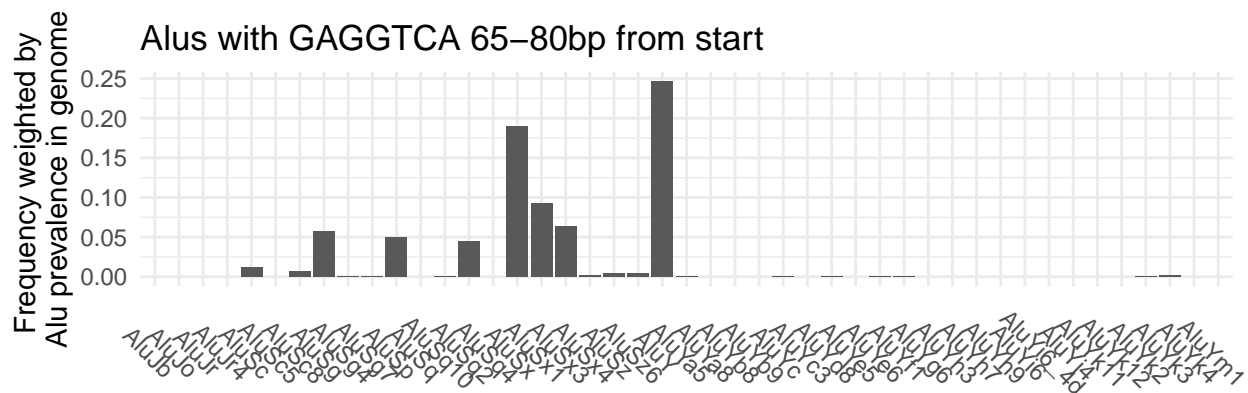
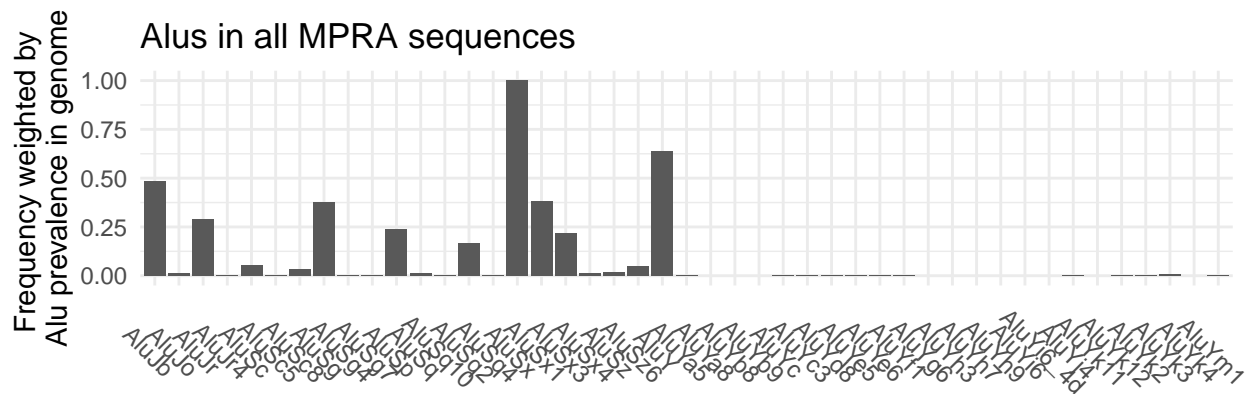
```
## pdf
## 2
```

Motif subtypes by distance

Are the repeat elements that have 'GAGGTCA' 65-80 bp from the start enriched for a certain family?

```
## Warning: Removed 11 rows containing missing values (position_stack).
```

```
## Warning: Removed 23 rows containing missing values (position_stack).
```



```
## Warning: Removed 11 rows containing missing values (position_stack).
```

```
## Warning: Removed 23 rows containing missing values (position_stack).
```

pdf

2

No.