Hypotheses

1. Transposon distribution affected by contact persistence

- high-cp contacts likely to be harder to penetrate due to frequent interaction of regions

Test

> Permutation test of repeat enrichmen1:

a. Repeats + Transposon – lower at persistent regions

b. Repeats

c. Transposons

d. 56 families

e. 372 subfamilies in age ranking

Check changes when using results without cp bed filtering based on foi chromosomes

Conclusion

> Certain subfamilies enriched and depleted at persistent contact regions

> Accounting for copynumber (insertion frequency) and considering subfamilies in age rank, enriched ones are dominantly old (mainly MIR) while depleted ones dominantly young (mainly Alus)

> Not so sure about Conclusion 2 cause only 2 families drive the trend, also L1 subfamilies with age shows the opposite trend (although expected from reports that Alu and L1 are distributed differently across the genome)

> Although we can focus on dominant families in the genome and their trends

2. Distribution of repeat sites between pairs of regions of high-cp contacts

- Could be due to frequent interaction of regions forming high-cp contacts

- Distribution could be via co-insertion or recombination

Test

> GC-skew repeat version

- per contact, absolute difference of repeat sites between regions divided by total repeat sites

- we don’t use the span of the repeat because insertion instance not defined by the span of the remnant

- the closer the value to 0, the more distributed the repeat sites are between regions in contact

- this metric will not suffice for hypothesis regarding repeat contribution to complementarity because it cannot for instance differentiate between 10 shared sites vs. 5 shared sites

- no need for length bias normalisation to compare results between families (because we don’t care about the absolute shared number of sites)

3. Repeats as additional reinforcers of complementarity

Test

> Shared number and fraction of contacts with at least 1 shared site

- shared number cannot be compared between families due to length bias (needs normalisation)

> Compare with shuffled?

- think of better metric to measure contribution to complementarity?

4. Complementarity not a repeat phenomenon

Test

> Complementarity after masking repeats + transposon, repeats only, transposon only

**Note**

1. Based on initial analysis, most contacts do not contain shared sites for most of the repeat families, so we’ll often get a GC-skew metric of 1

> Then characterise repeat families based on their behaviour per metric/test