Report 7: GAT tweaks

Cyril Matthey-Doret

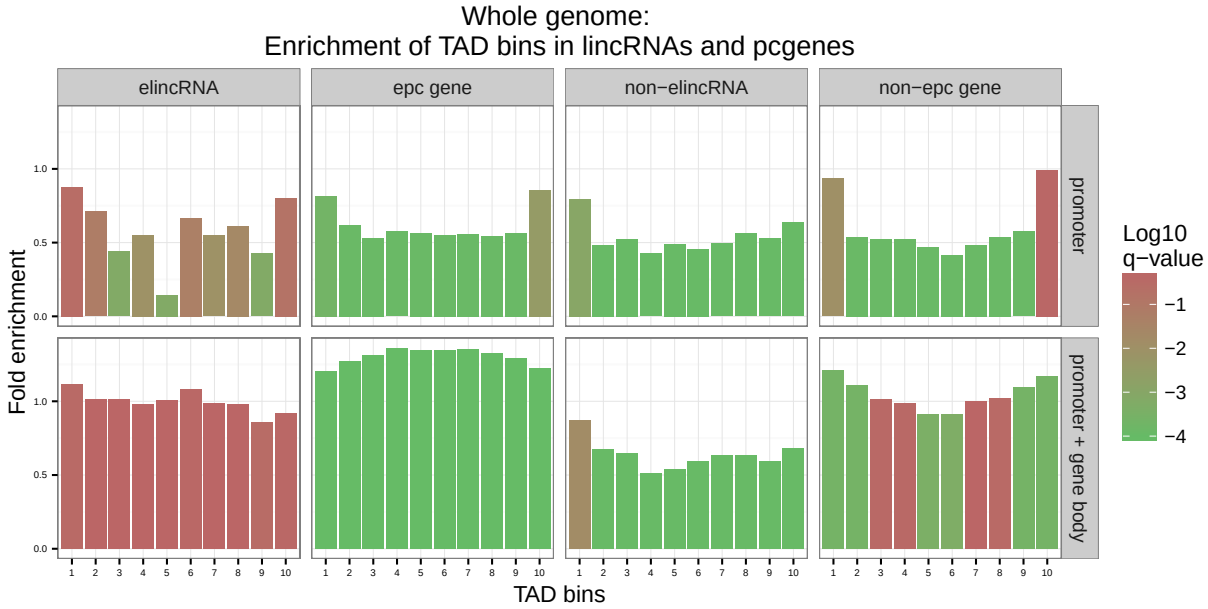
4 November 2016

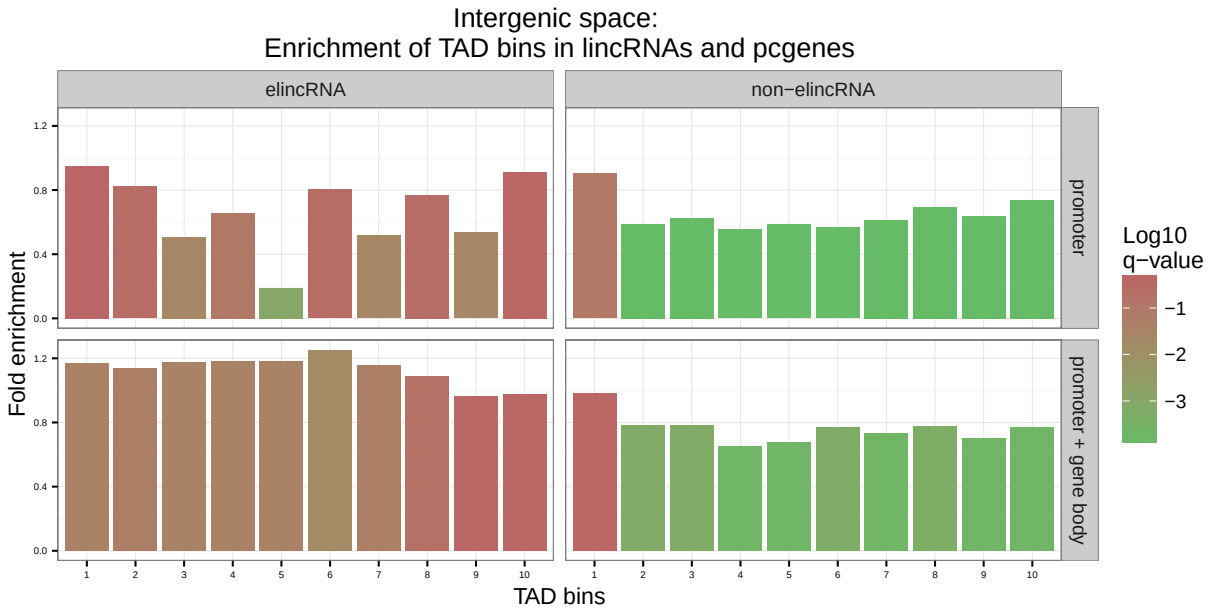
Introduction

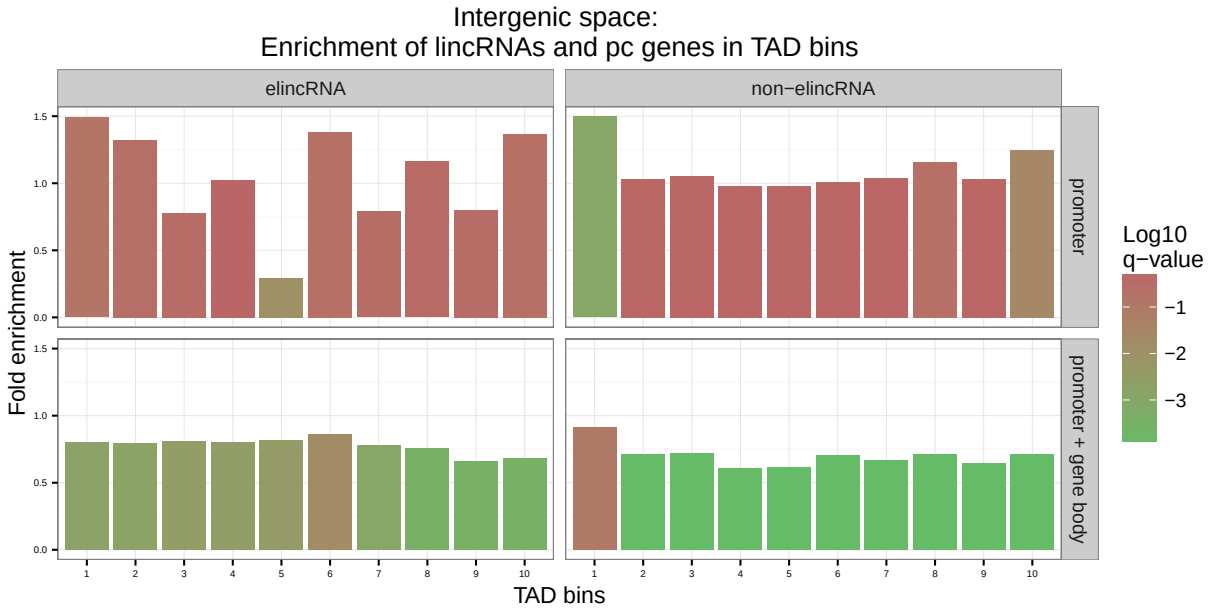
Here I increased the sampling of GAT to 10000 instead of 1000 to give it more power. I also worked with 10 bins per TAD instead of 20. I only used segment overlap as a measure, not nucleotide. When using expressed lincRNA, intergenic space and all protein-coding genes as workspace, I had to change the –nbuckets parameter from the default of 100000 to 270000 so that nbuckets*bucket size was larger than my largest segment of 269000. This also slows down the computations. Finally, I split the binding sites for CTCF, SCM3 and RAD21 into 3 different analysis so that if the results are only positive for 1 of the proteins, it will not be hidden.

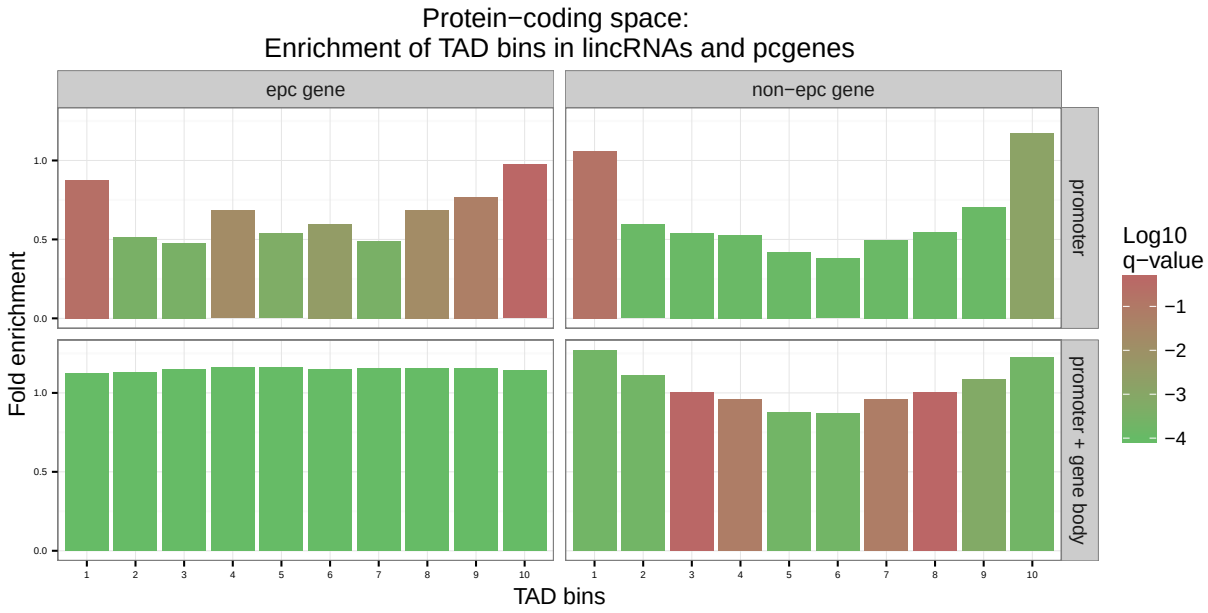
Results

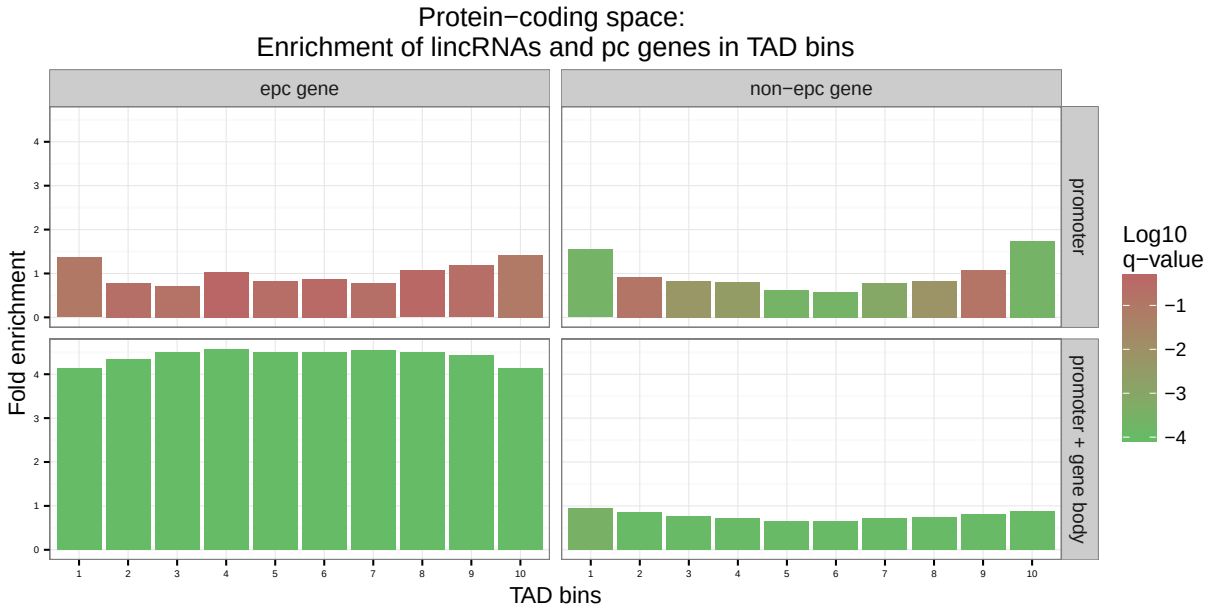
Analysis using TAD bins

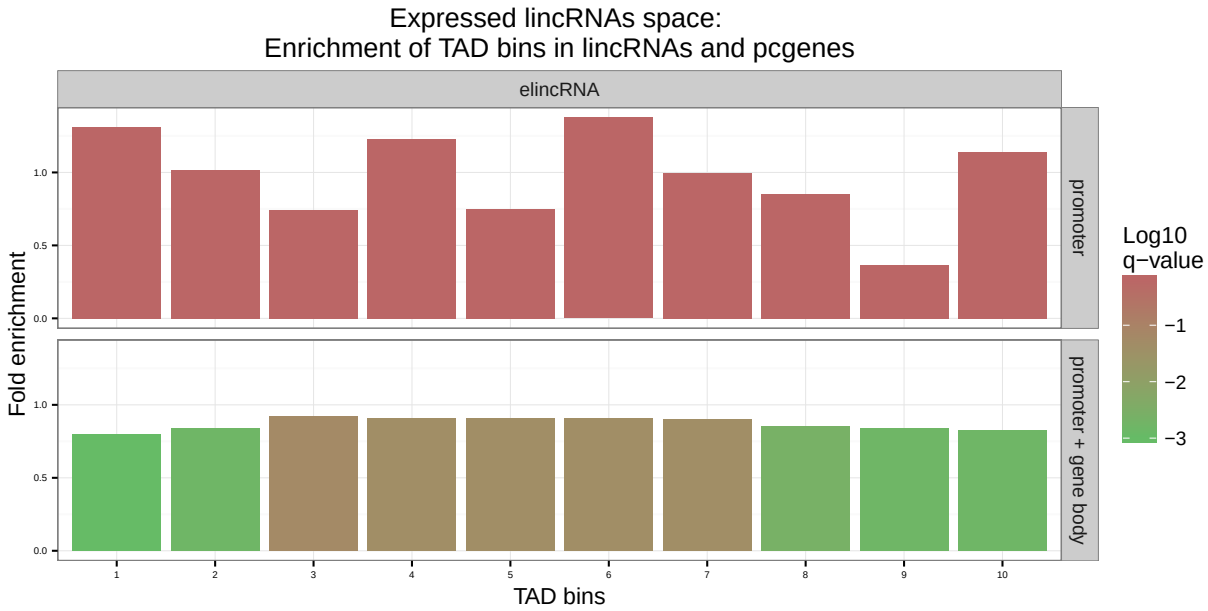



Intergenic space:
Enrichment of TAD bins in lincRNAs and pcgenes
elincRNA
non−elincRNA
promoter
promoter + gene body
Fold enrichment
TAD bins
Log10
q−value
−1
−2
−3


Intergenic space:
Enrichment of lincRNAs and pc genes in TAD bins
elincRNA
non−elincRNA
promoter
promoter + gene body
Fold enrichment
TAD bins
Log10
q−value
−1
−2
−3

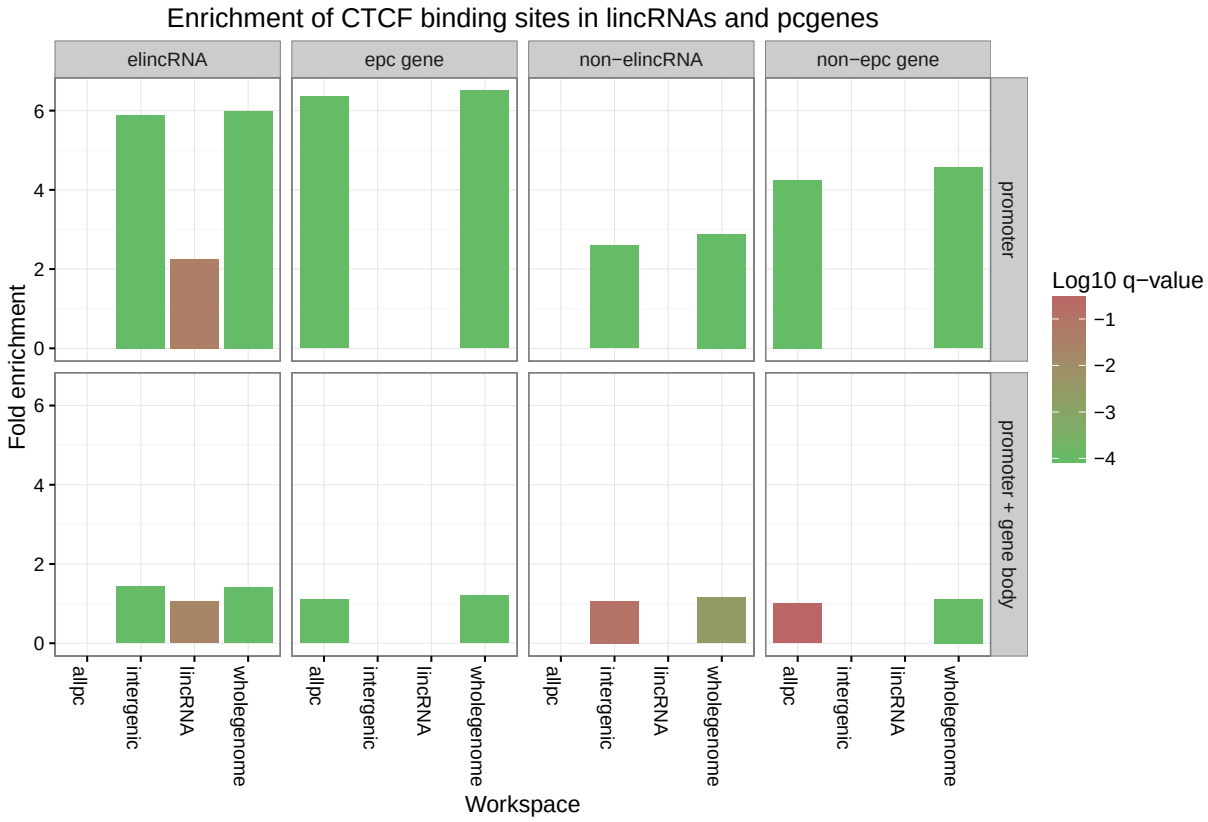

Protein−coding space:
Enrichment of TAD bins in lincRNAs and pcgenes
epc gene
non−epc gene
promoter
promoter + gene body
Fold enrichment
TAD bins
Log10
q−value
−1
−2
−3
−4

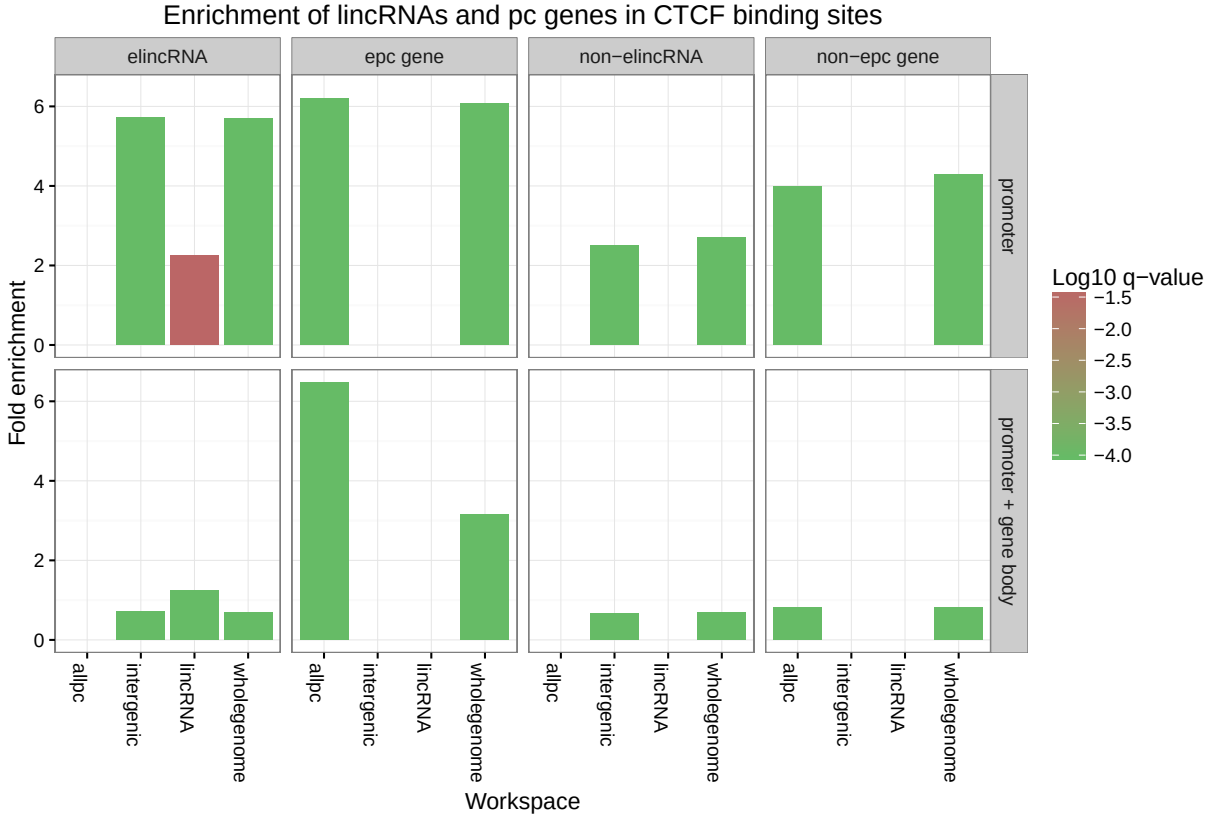

Protein−coding space:
Enrichment of lincRNAs and pc genes in TAD bins
epc gene
non−epc gene
promoter
promoter + gene body
Fold enrichment
TAD bins
Log10
q−value
−1
−2
−3
−4


Expressed lincRNAs space:
Enrichment of TAD bins in lincRNAs and pcgenes
elincRNA
promoter
promoter + gene body
Fold enrichment
TAD bins
Log10
q−value
−1
−2
−3

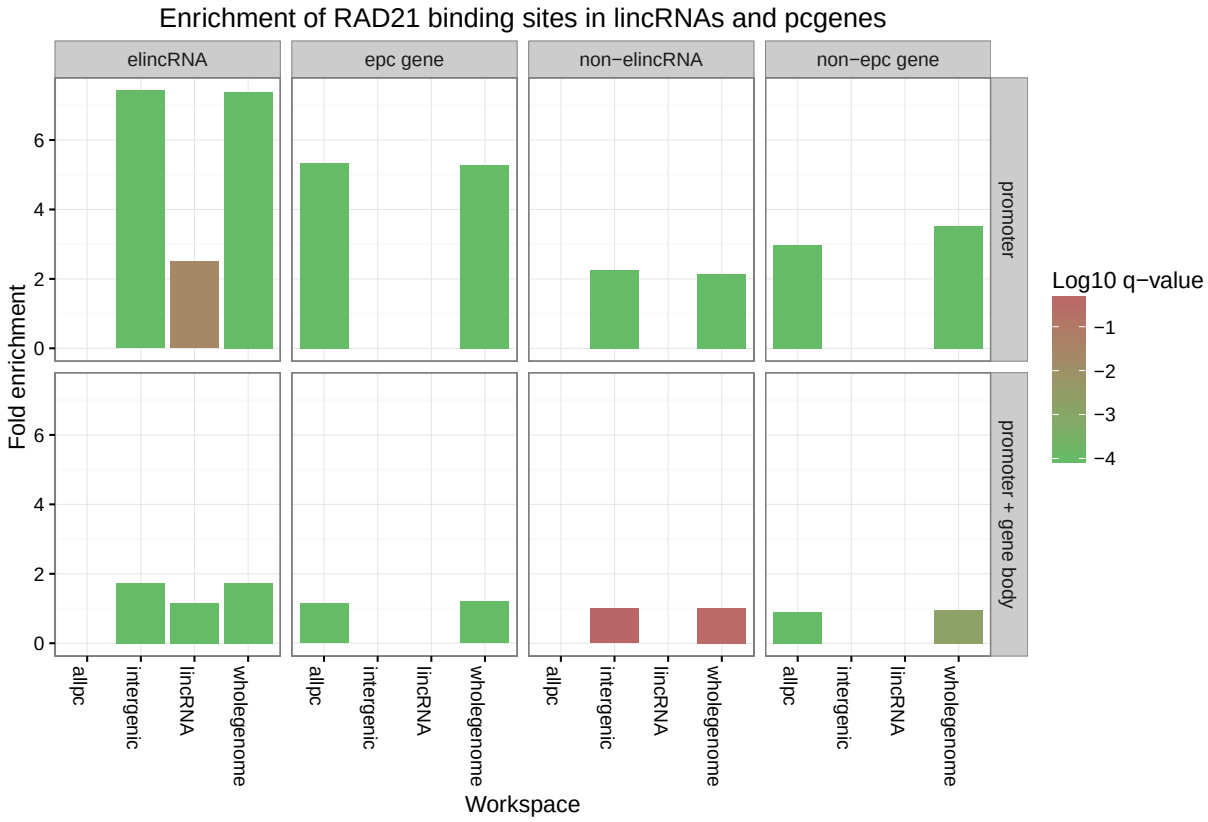

Expressed lincRNAs space:
Enrichment of lincRNAs and pc genes in TAD bins
elincRNA
promoter
promoter + gene body
Fold enrichment
TAD bins
Log10
q−value
−1
−2
−3
−4

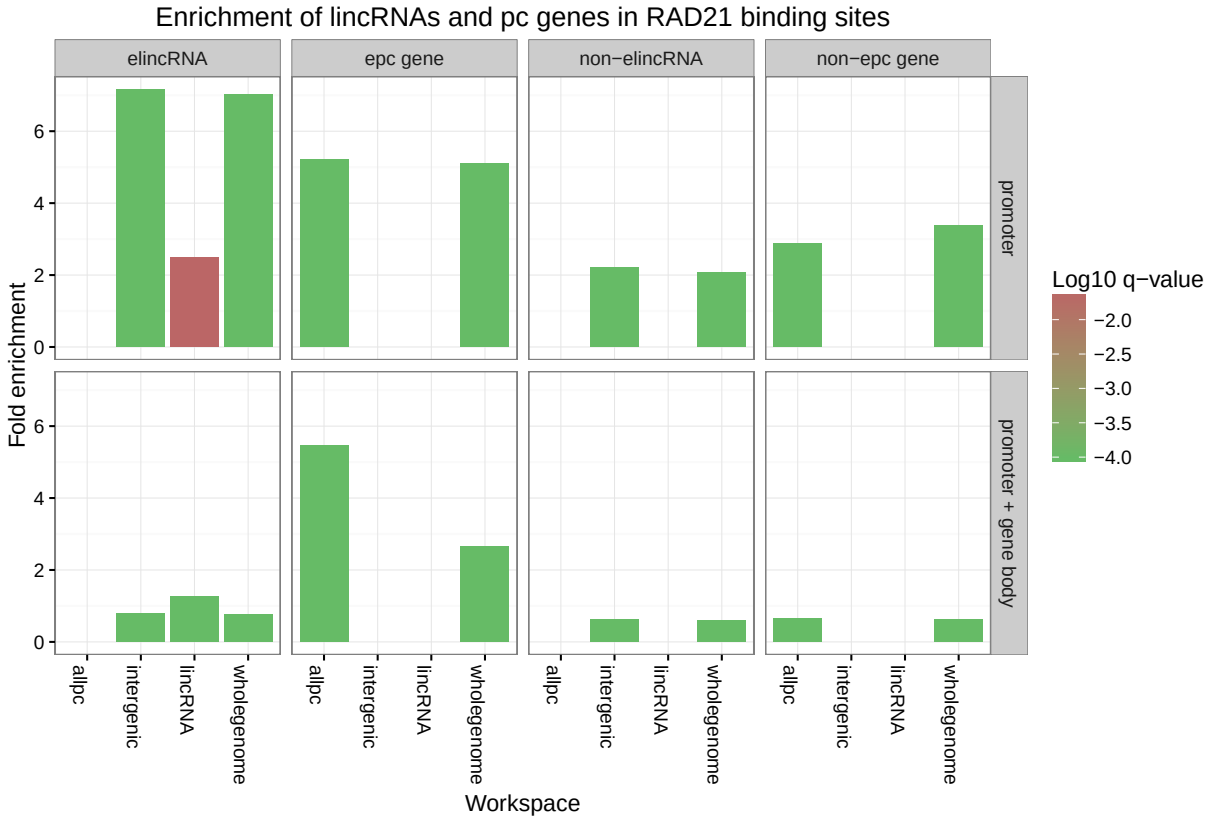
Analysis using CTCF binding sites



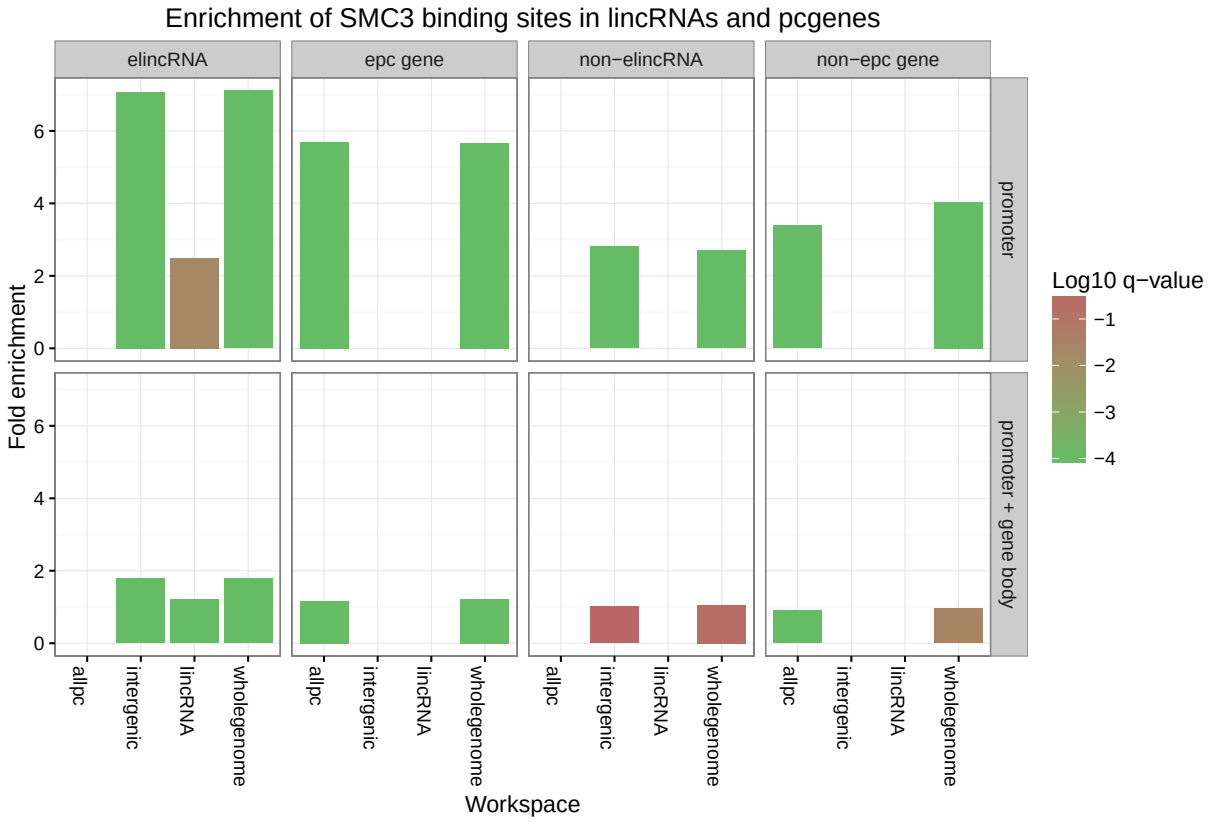


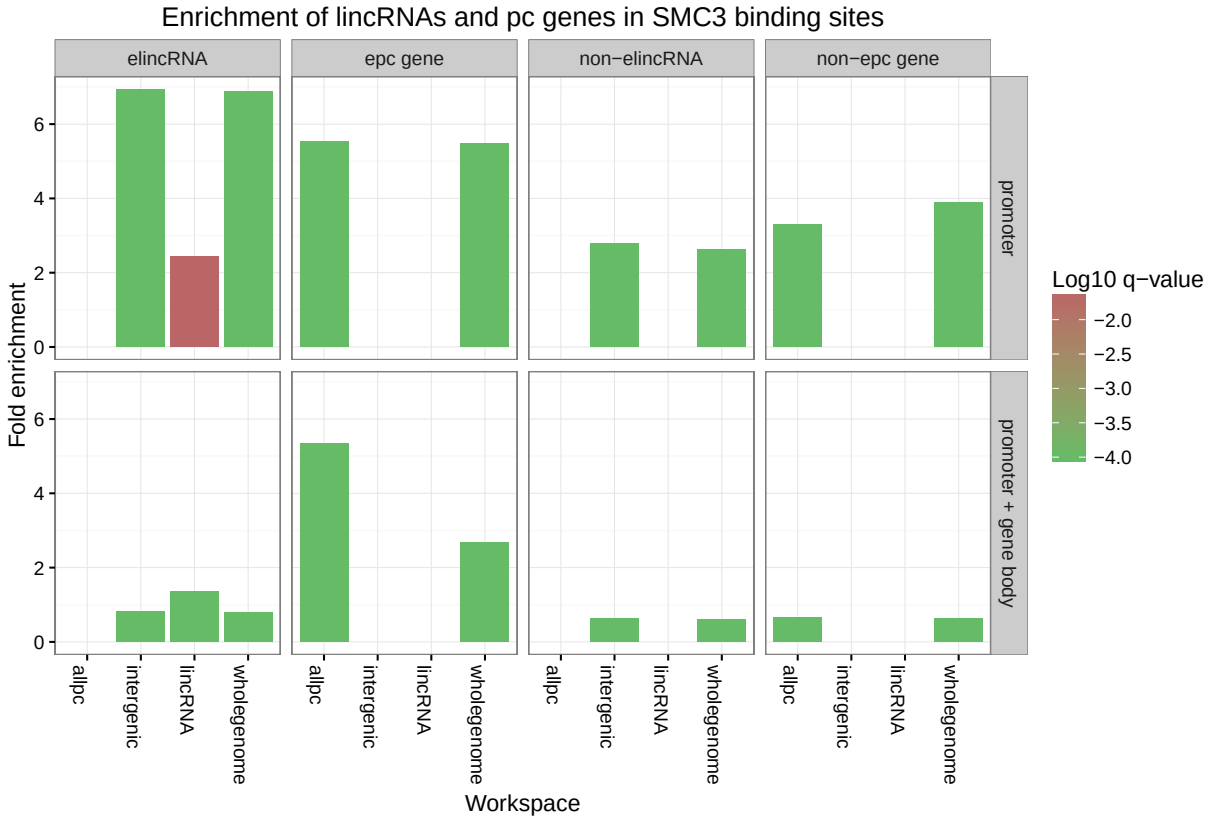
Analysis using RAD21 binding sites





Analysis using SMC3 binding sites





Conclusions

All promoters seem to be enriched at TAD borders, although the enrichment is not significant for elincRNA promoters. elincRNA promoters are strongly depleted in the middle of the TADs (bin 5). epc gene bodies are strongly enriched throughout TADs compared to all other genes types.

Results seem rather consistent across binding sites: elincRNA promoters are strongly enriched, followed by epc promoters. epc gene bodies are strongly enriched inside binding sites, but the inverse does not hold(enrichment of binding sites in epc gene bodies).