Report 2: Analyses of TADbound and non-TADbound lincRNAs after merging TADs

Cyril Matthey-Doret

October 11, 2016

TAD boundaries definition

As mentioned in the previous report, the original definition of TAD boundaries was problematic, because of the high number of overlapping TADs. Here, TAD boundaries are defined in the same way as before, except the TAD used are merged to consider only the large ones. By doing so, the number of TADs has been reduced from 9274 to 4533.

Overlapping genes with TAD boundaries

Both lincRNAs and protein coding genes have been overlapped with TAD boundaries, using the intersect program with -f 0.25 (genes must have at least 25% of their sequence overlapping a boundary to be considered). The set of overlaps was then cleaned by removing duplicates.

Data

lincRNA

Original set	Threshold	Overlaps	TAD-bound genes
2510	5%	493	419 (16.7%)
	10%	866	684 (27.3%)
	20%	1527	1121 (44.7%)

Protein-coding genes

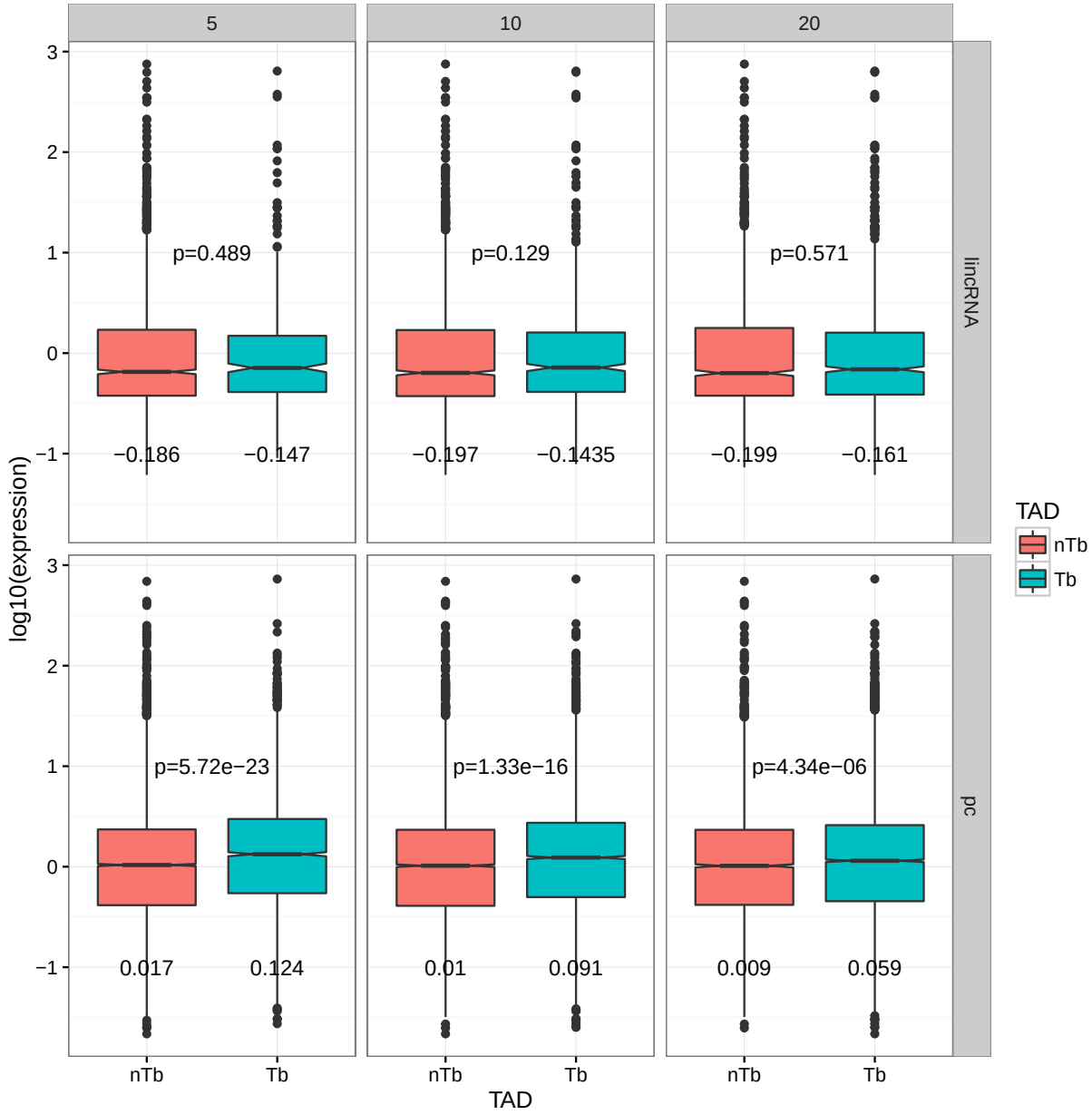
Original set	Threshold	Overlaps	TAD-bound genes
14846	5%	4020	3306 (22.3%)
	10%	7288	5575 (37.6%)
	20%	12873	8980 (60.5%)

Results

All p-values displayed on plots were obtained by performing Mann-Whitney tests.

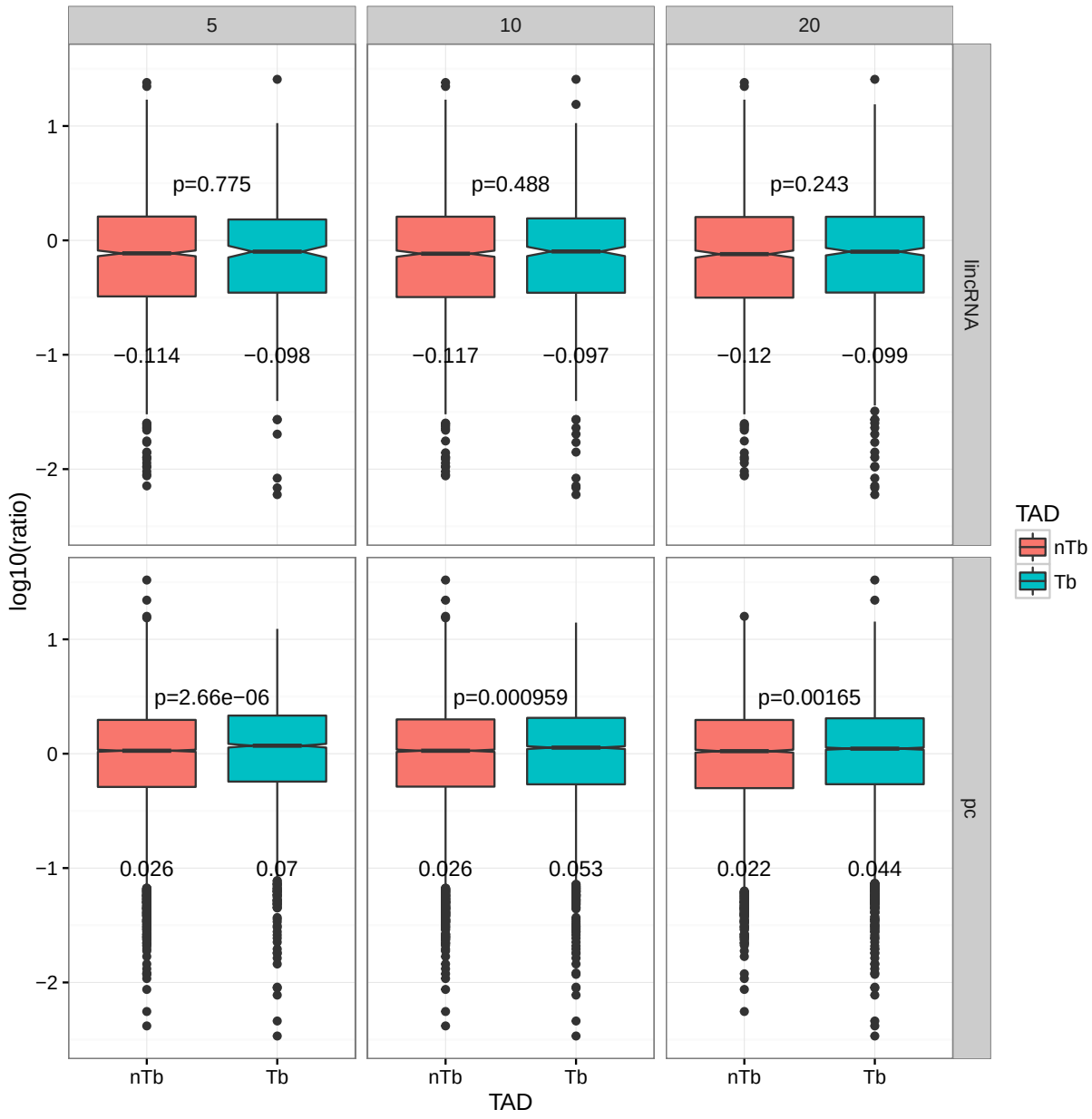
Expression levels

It seems the threshold used to define TAD-boundariesdoes not have much influence on the results. The TAD-bound lincRNAs have similar expression levels to the non-TAD-bound lincRNAs, regardless of which threshold was used, whereas the TAD-bound protein-coding genes where always more expressed than the rest.



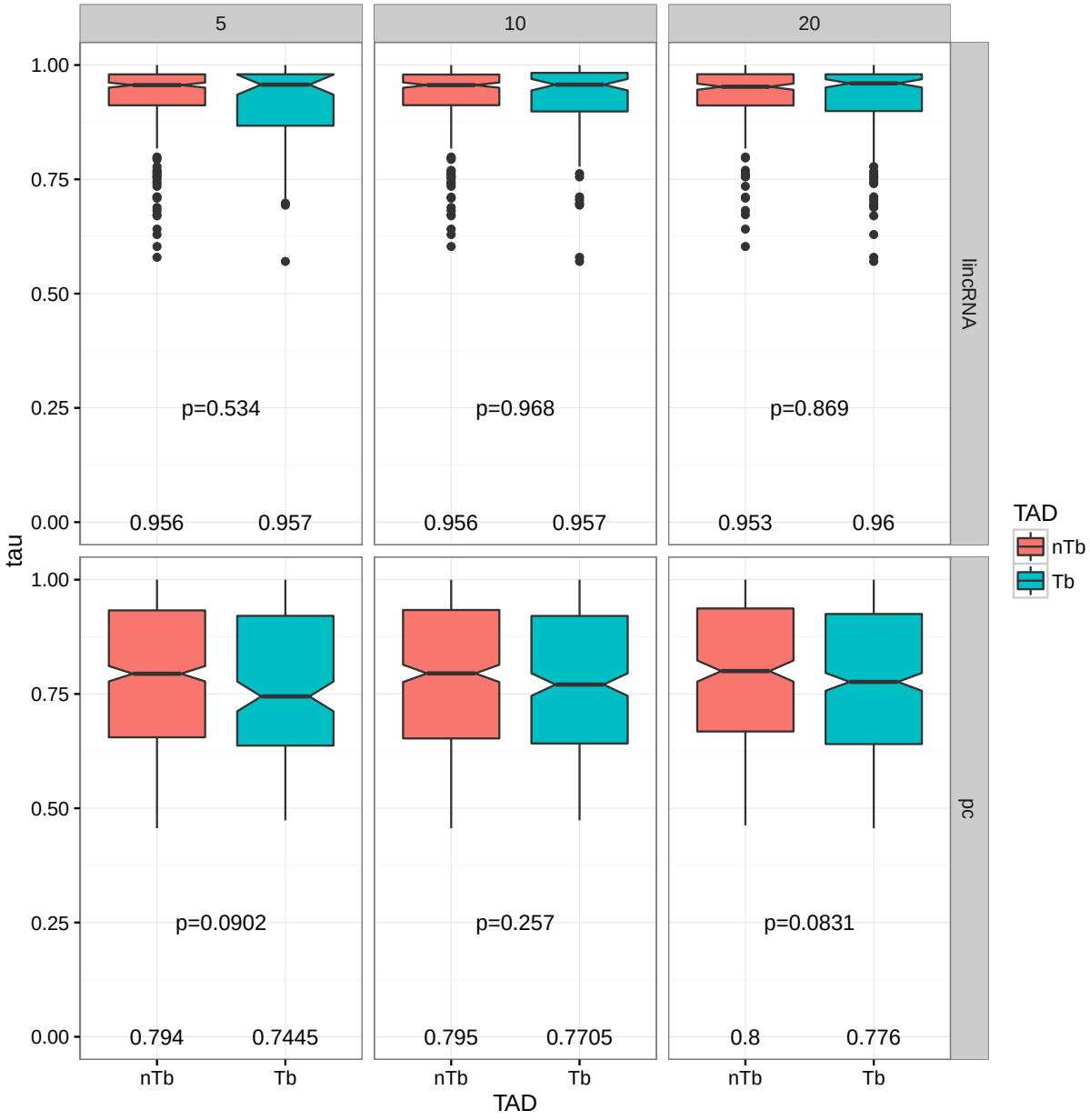
Subcellular localization

The ratios depicted below represent the abundance of RNA in the cytoplasm relative to the nucleus. Again, regardless of the threshold used to define TAD-boundaries, There is no significant difference between the localization of TAD-bound and non-TAD-bound lincRNAs, while the TAD-bound protein-coding genes are always more abundant in the cytoplasm than in the nucleus.



Tissue specificity

τ was calculated from a set of tissue-wise expression level for all transcripts. Neither lincRNA, nor protein coding gene show significant differences in tissue specificity between TADbound and non-TADbound genes. Overall the lincRNAs are more tissue specific than protein coding genes.



Conclusion

Merging the TADs does not seem to have much effect on the results. It would seem either that expression levels, tissue specificity and subcellular localization of lincRNAs are not affected by proximity to TAD boundaries, or that the method used to define TAD boundaries is too broad and hides differences.