

Report 1: Expression level, subcellular localization and tissue-specificity of TAD-bound and non-TAD-bound lincRNAs.

Cyril Matthey-Doret

October 6, 2016

TAD boundaries definition

When defining TAD boundaries as the areas 20% before and after TAD borders, those boundaries seemed to overlap with those of the neighbouring domains. To solve this issue, the TAD boundaries have been set to 3 different, arbitrary threshold: 5%, 10% and 20% of the TAD length.

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%	865	609 (24.3%)
	10%	1481	960 (38.2%)
	20%	2554	1420 (56.6%)

Protein-coding genes

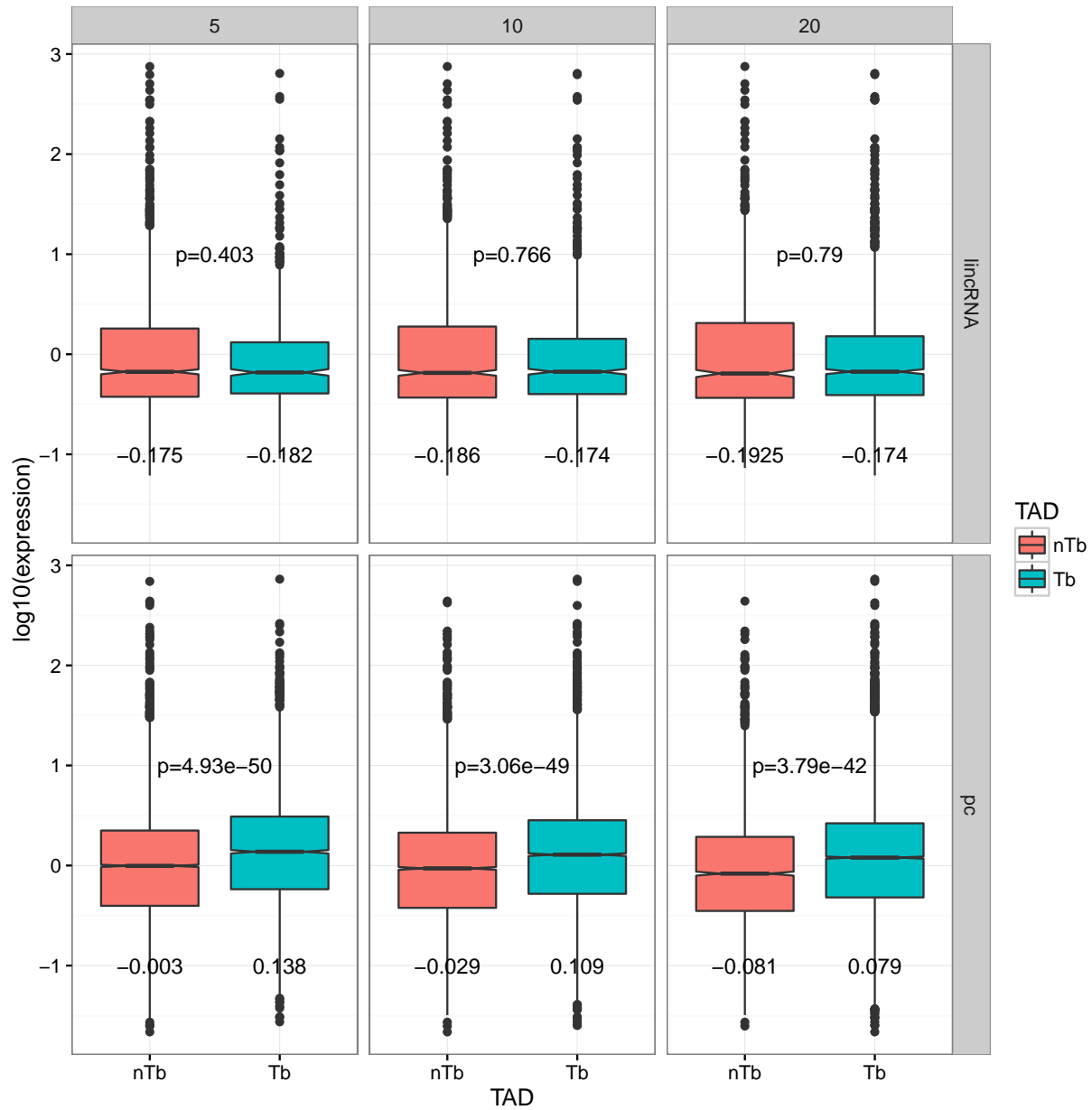
Original set	Threshold	Overlaps	TAD-bound genes
14846	5%	6576	4607 (31%)
	10%	12017	7495 (50.5%)
	20%	21544	11337 (76.4%)

Results

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

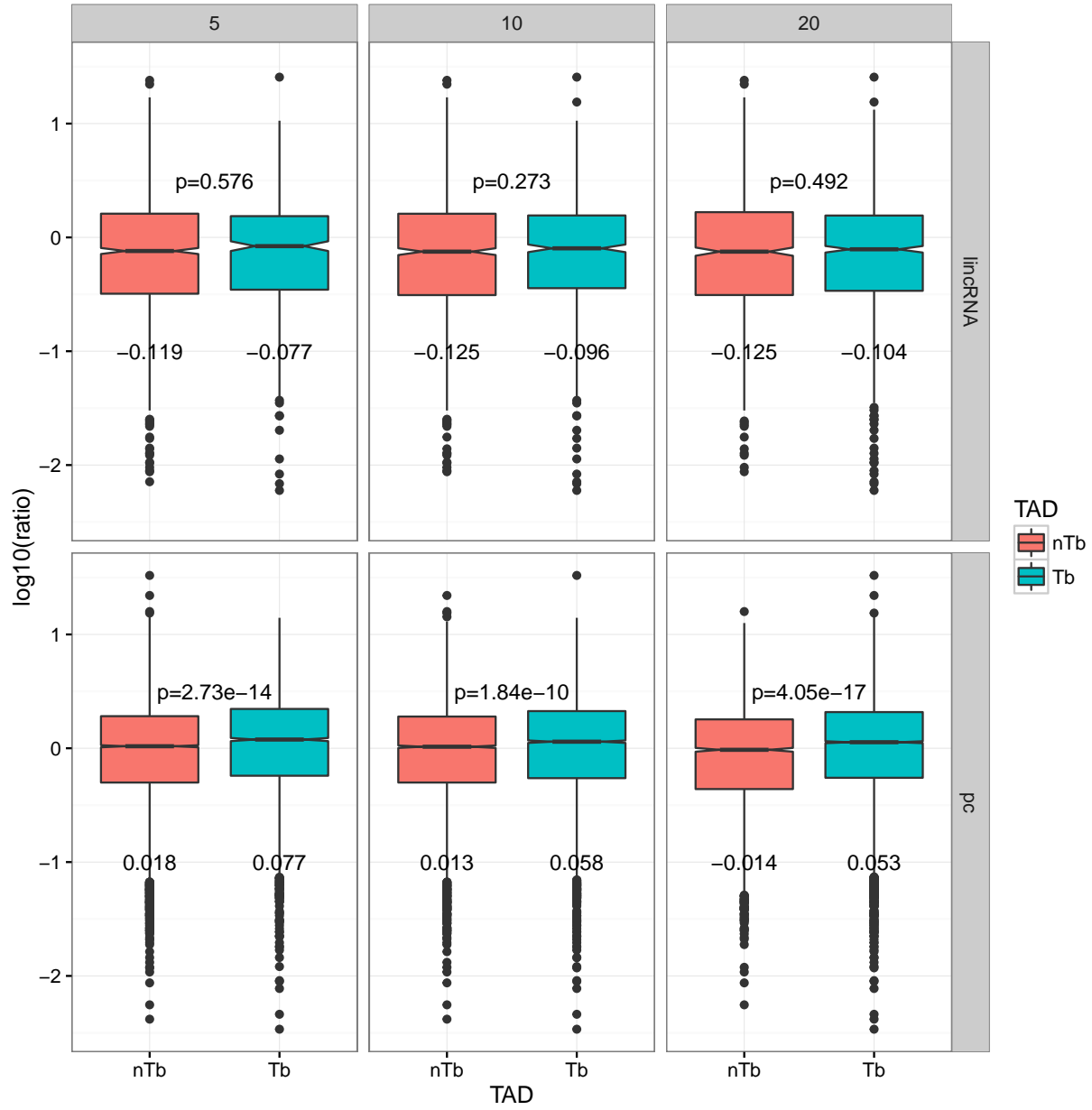
Expression levels

It seems the threshold used to define TAD-boundaries does 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 were always more expressed than the rest.



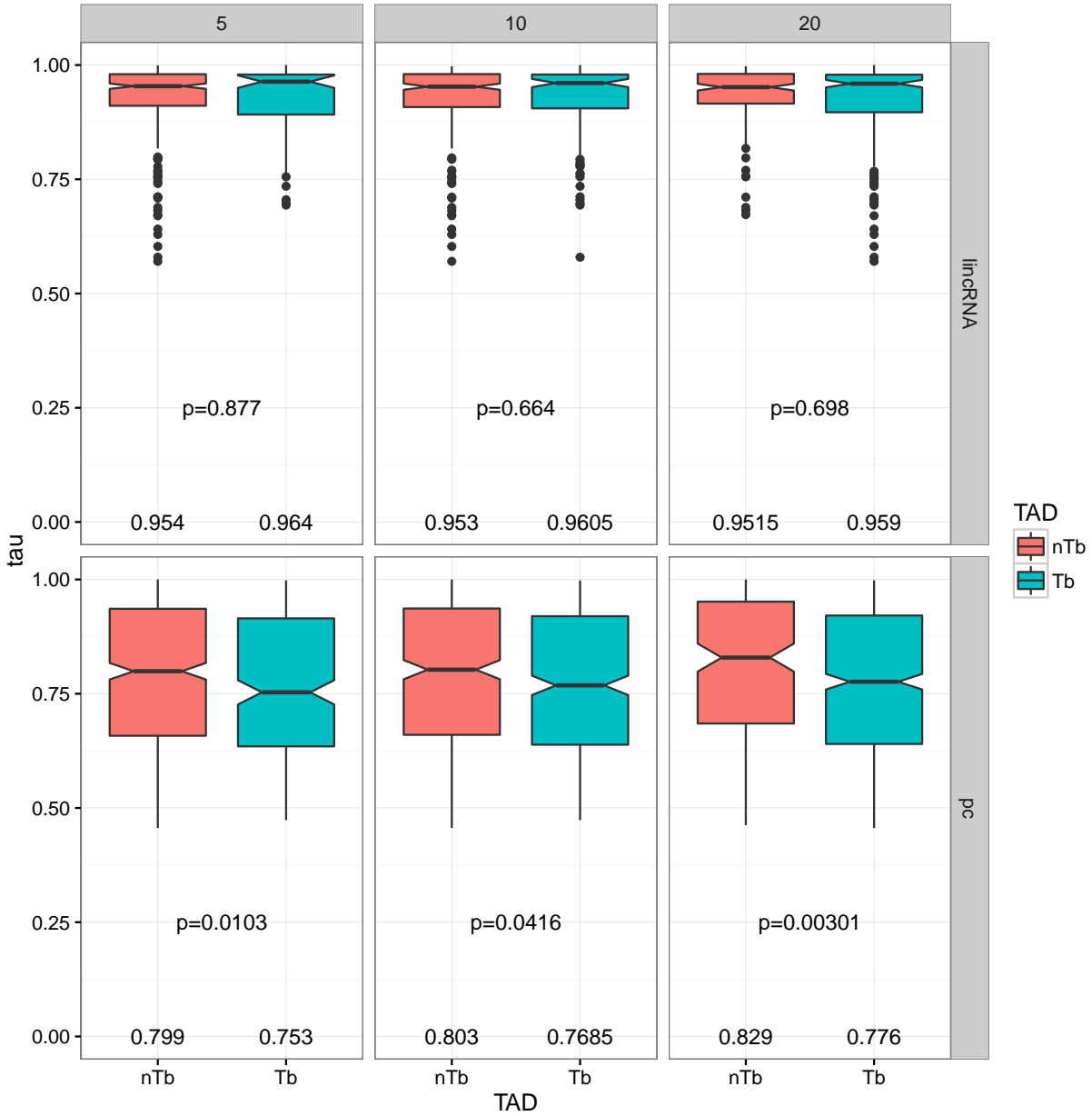
Subcellular localization

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. Again, there seem to be no difference in tissue specificity between TAD-bound and nonTAD-bound lincRNAs, whereas the TAD-bound protein coding genes appear to be less tissue specific. Overall the lincRNAs are more tissue specific than protein coding genes.



Conclusion

The TAD definition might not be appropriate. It was already pointed out (Rao et al, 2014) that the large TADs often contain many smaller TADs (see pictures from UCSC). These overlaps often cause the TAD boundaries to overlap and probably generate a lot of noise in the results. It might be more interesting to merge overlapping TADs in order to consider only the largest ones and their boundaries.