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

Cyril Matthey-Doret

October 25, 2016

TAD boundaries definition

This is the same as report 1, except TAD boundaries have been refined. They are calculated as 5% around TAD borders, and if a TAD should go further than half the gap until the next TAD, it is instead resized to this value (half the gap). Note TADs have been merged to remove the small ones, and boundaries are based on the resulting large TADs.

Overlapping genes with TAD boundaries

Two different thresholds were chosen to consider overlaps. lincRNAs needed at least 51% overlap, while protein-coding genes needed 11%. Those thresholds were chosen based on the mean value of median overlap length for both lincRNAs and protein-coding genes.

Data

lincRNA

Original set	Threshold	Overlaps	TAD-bound genes
2510	5%	437	375 (14.9%)

Protein-coding genes

Original set	Threshold	Overlaps	TAD-bound genes
14846	5%	5687	4534 (30.5%)

Results

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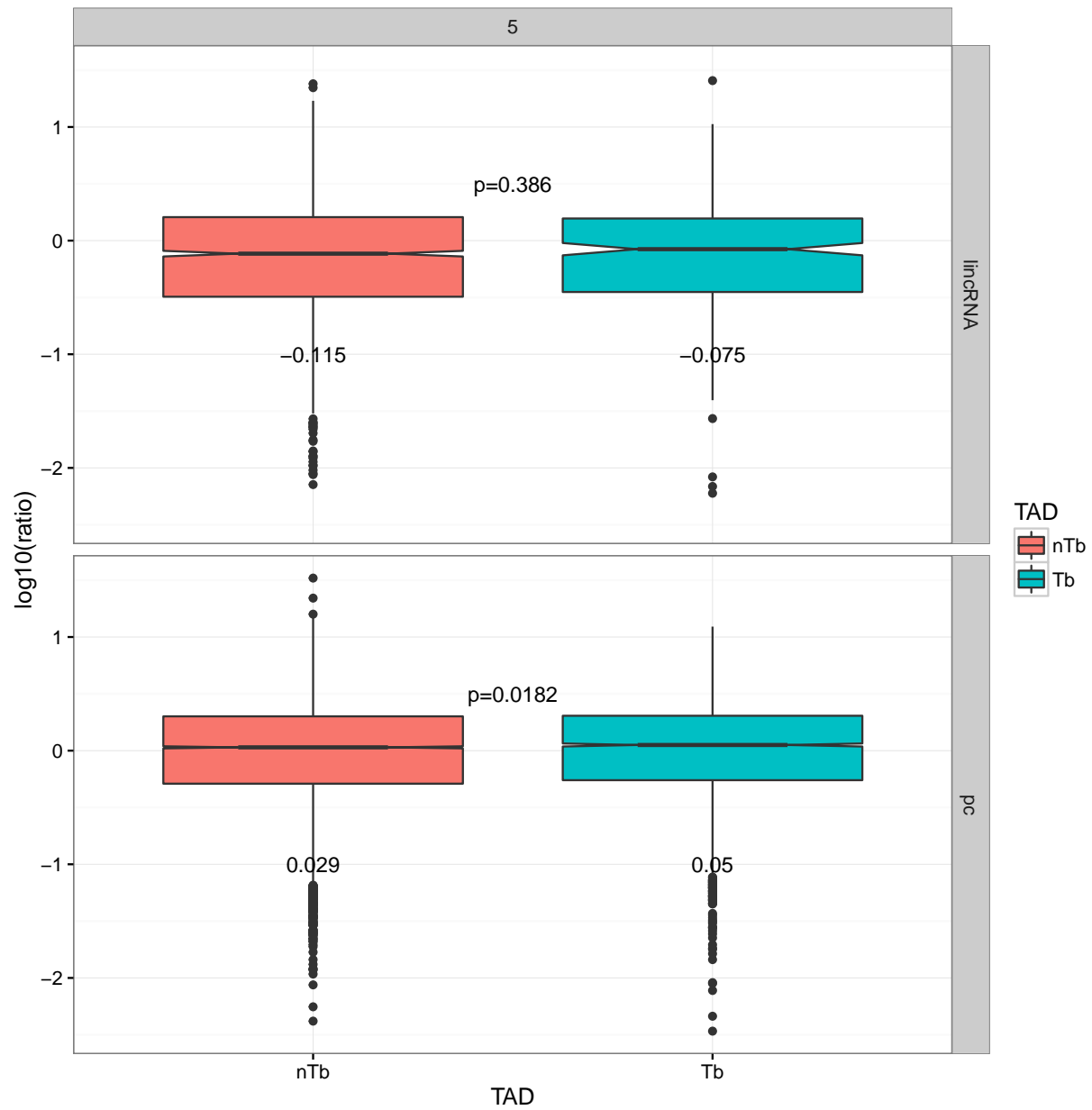
Expression levels

TAD-bound protein-coding genes where always more expressed, but there is no difference for lincRNAs.
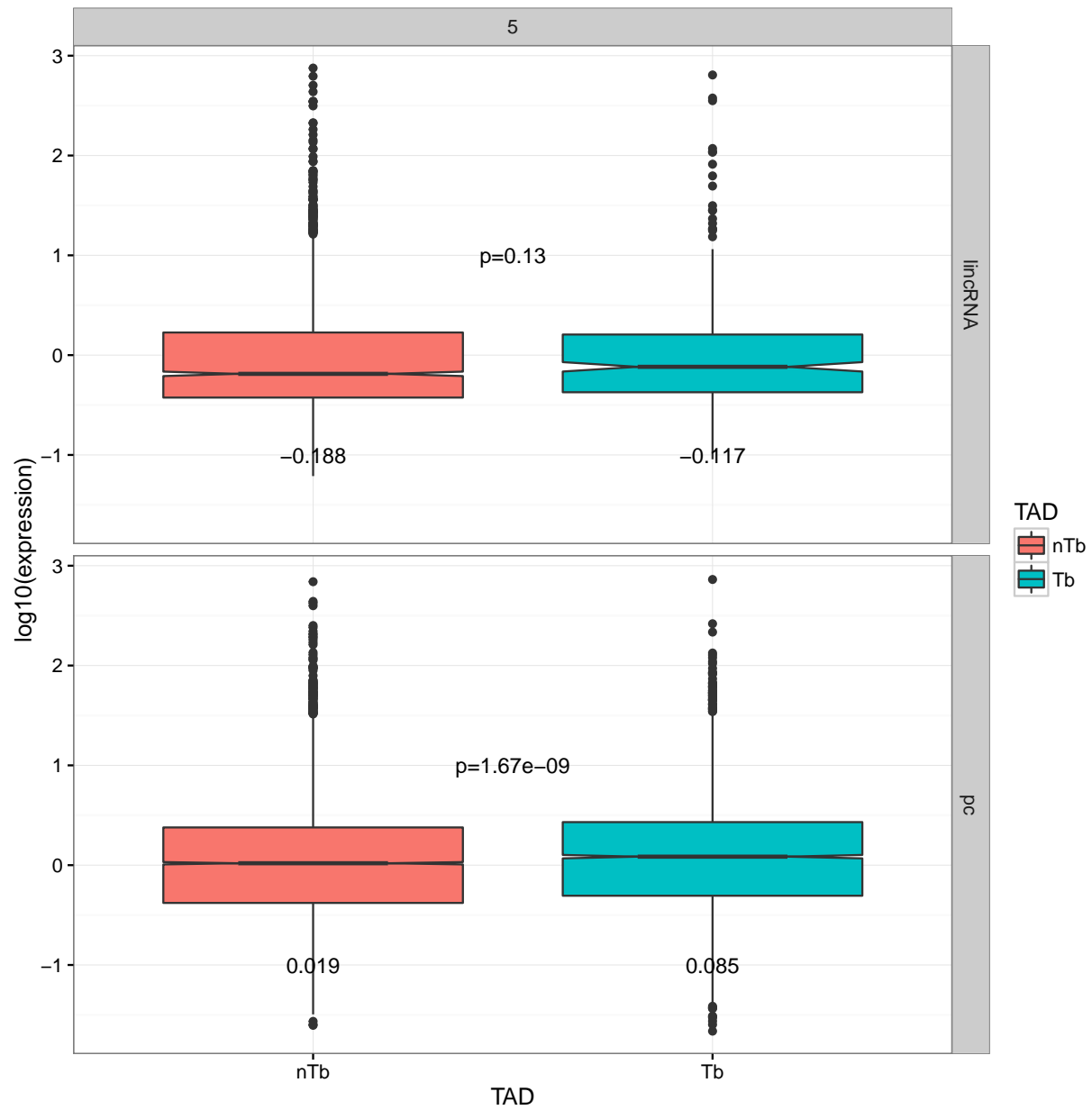
Subcellular localization

The ratios depicted below represent the abundance of RNA in the cytoplasm relative to the nucleus. There is no significant difference between the localization of TAD-bound and non-TAD-bound lincRNAs, while the TAD-bound protein-coding genes are 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 or protein coding genes. Overall the lincRNAs are more tissue specific than protein coding genes.

```
## notch went outside hinges. Try setting notch=FALSE.
```
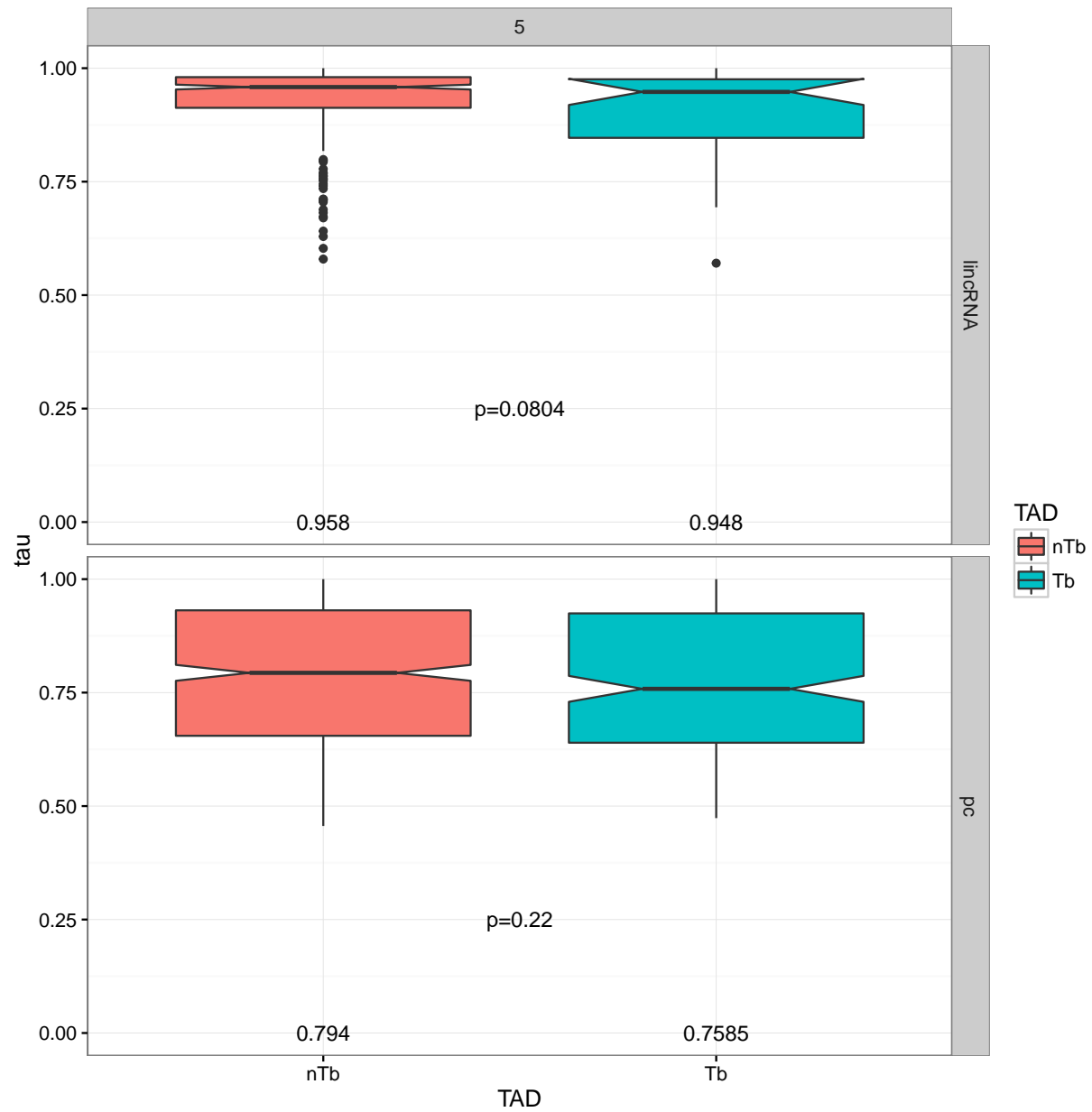
Conclusion

The TAD definition is still not appropriate. Giving up this track to define TAD boundaries. A more meaningful way would be to use raw-hic data and define boundaries based on the number of interactions observed.