Report 3: Redefining boundaries

Cyril Matthey-Doret

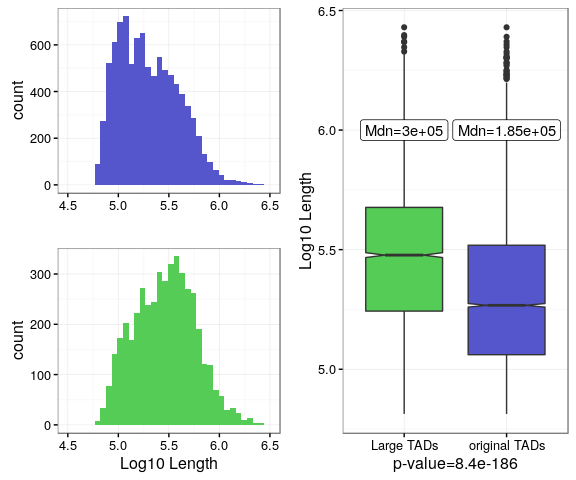
11 octobre 2016

# Introduction

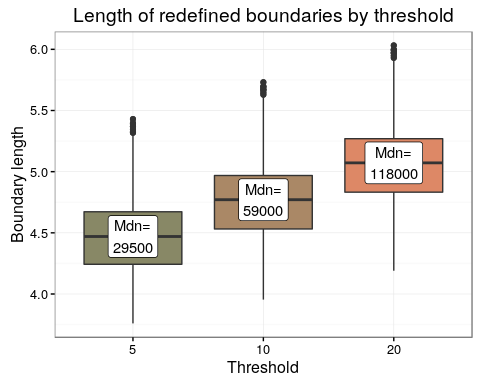
Visualising and quantifying the distribution of TAD lengths, or the areas of ovelap between lincRNAs and TADs might help finding a better threshold for the definition of TAD boundaries. This report aims to provide support for the definition of those boundaries.

# 1. Length distribution of TAD, boundaries and gaps

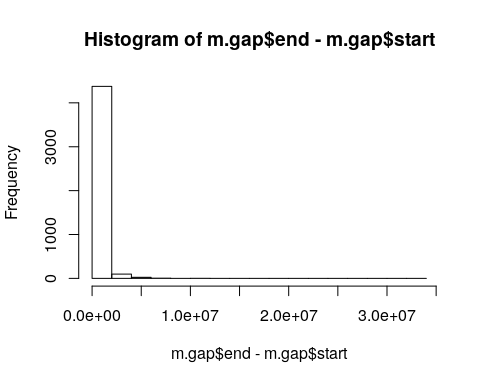
* Length of TADs: TADs often contained many smaller TADs which causes the boundaries to overlap and makes it harder to distinguish between TAD-bound and non-TAD-bound RNAs.



* Length of TAD boundaries: TAD boundaries have been redefined so that they are more flexible. Again, three different thresholds have been used to define them, however a boundary cannot overlap another. A TAD boundary will always reach only as far as the smallest value between the threshold and half the gap between its TAD and the neighbour.



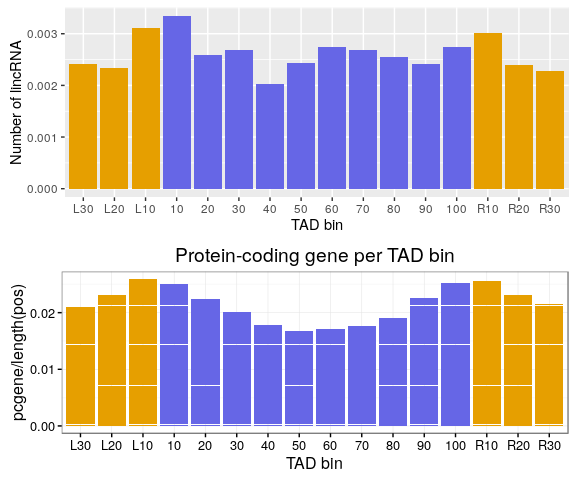
* Length of gaps: Distribution gap length.



# 2. Overlaps between RNA and areas of TAD boundaries

TADs were divided into 10 bins, plus 3 outer bins on both sides. The length of each bin is 10% of TAD length. All RNAs with at least 25% of their sequence overlapping a TAD bin. This allows to in which area of the TADs linc RNA and protein coding genes are more likely to be expressed. The barplots below display the total number of lincRNAs or protein-coding genes that overlap with each bin. The bins depicted in orange are located outside TADs, while the ones in blue are inside. All bins beginning by L are on the left (before the TAD), while those beginning with R are on the right (after the TAD).

## NULL

 #3. ...