Tale on the transposons on chromatin landscape

First Author, V. N. Babenko

Institute of Cytology and Genetics, Lavrentieva str., 10, Novosibirsk, 630090 Russia, bob@bionet.nsc.ru

V. F. Matvienko

Novosibirsk State University, Koptyuga 2, Novosibirsk, 630090 Russia, vmatv@ngs.ru

We categorized human genome 100kb non-overlapping segments by their Dnase Hypersensitive Sites (DHS) counts based on data in (Sheffield et al., 2011). They fit a Weibull long tail distribution with a peak at around 14 DHSs per bin. The few (around 50) bins maintaining less than 14 DHSs were mostly gene deserts, long introns, or some quite distinct gene clusters like ubiquitin peptidase family.

Then we performed linear regression analysis between categorized by families transposons counts and #DHS. We revealed two major classes of transposons families: those that prefer "silent" chromatin and those tending to reside in "open" chromatin bins with high confidence. Further on, we discovered that number of Alu retroposons strongly correlates with the number of genes in the bin. Based on this observation, we worked out a method based on a non-linear Alu-gene correlation to infer some non-linear evolution events like the emergence of tandem repeated gene clusters.

We also crossed the family – categorized transposons with Txn table (transcription factor binding sites verified by Chip-Seq; genome.ucsc.edu) to elucidate their transposon specific propagation similar to (Jjingo et al., 2014). Further on, we assessed chromosome wise bias of repeat families and found that most chromosome – specific repeat families expansions (LINEs in majority) are maintained at X chromosome. Some ctcf-related open chromatin LTR expansions were observed specifically at chromosome 19 in a way similar to B2 Sine in mouse (Lunyak et al., 2007).

Overall we report that the properties of transposons distribution and density within a genomic segment can disclose its specific evolutionary history and features.

References

Acknowledgements can be given at the end (as a separate untitled paragraph). The style for references is given below (preferably with titles, if the space allows).

1. Jjingo D et al. Mammalian-wide interspersed repeat (MIR)-derived enhancers and the regulation of human gene expression. Mob DNA. 2014;5:14
2. Lunyak VV et al. Developmentally regulated activation of a SINE B2 repeat as a domain boundary in organogenesis. Science. 2007; 317(5835):248-251.
3. Sheffield NC et al. Patterns of regulatory activity across diverse human cell types predict tissue identity, transcription factor binding, and long-range interactions. Genome Res. 2013; 23(5):777-78