Review existing methods for Hi–C data analysis

H.L

April 10, 2014

After a literature review, I collected the following reports describing methods of Hi–C data analysis:

1. Peng, C. et al. The sequencing bias relaxed characteristics of Hi-C derived data and implications for chromatin 3D modeling. Nucleic Acids Research 41, e183e183 (2013).

2. Zhang, Z., Li, G., Toh, K.-C. & Sung, W.-K. 3D Chromosome Modeling with Semi-Definite Programming and Hi-C Data. Journal of Computational Biology 20, 831-846 (2013).

3. Ay, F., Bailey, T. L. & Noble, W. S. Statistical confidence estimation for Hi-C data reveals regulatory chromatin contacts. Genome Research (2014). doi:10.1101/gr.160374.113

4. Lu, Y., Zhou, Y. & Tian, W. Combining Hi-C data with phylogenetic correlation to predict the target genes of distal regulatory elements in human genome. Nucleic Acids Research 41, 10391-10402 (2013).

5. Hu, M., Deng, K., Qin, Z. & Liu, J. S. Understanding spatial organizations of chromosomes via statistical analysis of Hi-C data. Quant Biol 1, 156-174 (2013).

6. Imakaev, M. et al. Iterative correction of Hi-C data reveals hallmarks of chromosome organization. Nat Meth 9, 999-1003 (2012).

7. Yaffe, E. & Tanay, A. Probabilistic modeling of Hi-C contact maps eliminates systematic biases to characterize global chromosomal architecture. Nature Genetics 43, 1059-1065 (2011).

8. Hu, M. et al. HiCNorm: removing biases in Hi-C data via Poisson regression. Bioinformatics 28, 3131-3133 (2012).

9. Jin, F. et al. A high-resolution map of the three-dimensional chromatin interactome in human cells. Nature 503, 290-294 (2013).

Questions to answer

Idea 1 Gene content per chromosome (in terms of active genes) and gene interactions. (You have more *cis* than *trans* interactions).

Idea 2 To challenge the so-called "transcription factories". If "transcription factories" exist, given that the vast majority of 3D chromosomal interactions are *cis* interactions, the co-regulated genes should tend to be on the same chromosome. While this could be true in certain cases, I am pretty sure that there are genes on different chromosomes that are regulated by the same factor. Also mobility of chromatin in the nucleus is rather limited (see Spector and also Soutoglou publications). The "transcription factory" model could potentially apply in certain cases of master regulators (KLF-1 for example) (see Schoenfelder paper) but it is probably not the rule.

Idea 3 Mentioned by Misteli [1]. Hi-C data refer to populations. This may obscure the relationship of 3D chromatin structure and function. Even in the same population of cells, there are cells expressing a gene and others that do not (Mention the recent Spector paper on monoallelic expression here). What is the difference in the chromatin structure though when certain genes are expressed vs. when they are not? The recent development of single-cell Hi-C (if it really works as advertised) [2] may help answer this question.

Idea 4 The "kill many birds with one stone" hypothesis. If we suppose that co-regulated genes (by master transcription factors) are truly in close proximity in 3D space, a favorite and very economical –in terms of changes required– strategy in cancer would be to create master regulator behavior where it does not exist and mess up many genes all together. This could be done by fusing the activation-domain of a master regulator with the binding domain of an abundant transcription factor.

References

[1] Tom Misteli. Parallel genome universes. *Nature Reviews Genetics*, 30(1):55–56, January 2012.

[2] Takashi Nagano, Yaniv Lubling, Tim J Stevens, Stefan Schoenfelder, Eitan Yaffe, Wendy Dean, Ernest D Laue, Amos Tanay, and Peter Fraser.

Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature*, 502(7469):59–64, 2013.