Do enhancer-associated lincRNAs contribute to chromosomal organization ?

First Step Project
Molecular Life Sciences, Bioinformatics

Cyril Matthey-Doret
Supervised by: Jennifer Yihong Tan
Directed by: Ana Claudia Marques

Department of Computational Biology
Department of Physiology
University of Lausanne - Switzerland

December 16, 2016

Abstract

Introduction

It was only recently discovered that a surprisingly large proportion of the mammalian transcriptome does not code for proteins. To date, the number of annotated noncoding genes longer than 200 nucleotides (long noncoding RNA, lncRNA) excess by at least 3 times that of protein-coding genes [Iyer et al., 2015]. Among lncRNAs, those that do not overlap with protein-coding genes are the most abundant (long intergenic noncoding RNAs, lincRNAs). Functional and evolutionary analyses, together with extensive characterization of a handful of lincRNAs, demonstrate that these transcripts are involved in gene regulation processes transcriptionally and post-transcriptionally, and that they can contribute to organismal traits and diseases [Kornienko et al., 2013]. However, the mechanisms and functions, if any, for the majority of lincRNAs remain unknown [Rinn and Chang, 2012].

LincRNAs associated with human traits have been shown to have enhancer-associated cis-regulatory roles and their loci are correlated with more compact chromatin relative to other lincR-NAs in a human lymphoblastoid cell line (LCL) (Tan et al, under revision). Most active enhancers are transcribed, generating noncoding products, including lincRNAs [Guil and Esteller, 2012]. This raises the question whether lincRNAs with enhancer-like activities (elincRNAs) contribute to gene regulation and the organization of chromosomal contacts. Unlike most enhancer-associated noncoding RNAs, which are often transcribed bidirectionally and then rapidly degraded [Darrow and Chadwick, 2013], elincRNAs are transcribed preferentially in one direction and are more stable [Marques et al., 2013]. These distinct features make them less likely to be a product of pervasive transcription and thereby, good candidates to study the involvement of lincRNAs in the regulation of gene-enhancer interactions within chromatin domains. Those elincRNAs will therefore be the focus of my analysis.

Recently, there have been reports of elincRNAs involved in the spatial organization of the genome, such as Haunt [Yin et al., 2015], which can regulate intrachromosomal interactions by mediating promoter-enhancer looping. It is thought that the architecture of the genome is an major factor in gene regulation [Engreitz et al., 2016]. Indeed, genomic DNA is folded into variably compact chromosomal structures that likely impact expression of the embedded genes [Gorkin et al., 2014]. On a global scale, regions with a high degree of compaction are classified as heterochromatin while relatively uncondensed regions are called euchromatin [Passarge, 1979]. These are respectively associated with lower and higher levels of active transcription [Tamaru, 2010]. Chromosomes are further compartmentalized into smaller domains, called topologically associated domains (TADs). The amount of DNA-DNA interactions is high within TADs as a result of their close spatial proximity, and low across different TADs. TAD boundaries are the regions lying at the borders of TADs (Figure 9) and are essential for gene regulation. They are often gene-dense and are enriched in highly transcribed genes [Ong and Corces, 2014].

Chromosomal contacts within TADs, often seen as looping structures, occur particularly at TAD boundaries and are crucial for establishing correct interactions between regulatory elements, such as enhancers and promoters [Gorkin et al., 2014]. Deletion of TAD boundaries often disrupts those interactions, resulting in gene misexpression and disease phenotypes [Lupiáñez et al., 2016]. TAD boundaries are also enriched in architectural proteins, including CTCF [Pope et al., 2014], which functions to delimit TAD borders by acting as genomic insulators that prevent DNA-DNA

interactions across multiple TADs. Cohesin, another architectural protein, is also enriched at TAD boundaries. It is a multi-protein complex that is thought to be involved in establishing enhancer-promoter interactions [Ji et al., 2016]. While most CTCF sites are shared between different cell types and species [Ji et al., 2016], cohesin binding at gene regulatory elements is often cell-type specific [Hadjur et al., 2009].

Using various bioinformatics tools to analyze publicly available multi-omics data from the ENCODE project [ENCODE Project et al., 2012] and data from whole-genome chromosome conformation capture (Hi-C) experiments [Rao et al., 2014], I investigated the molecular properties of elincRNAs, their enrichment in different regulatory regions and their association with the amount of DNA-DNA interactions to gain insight into their roles in gene regulation within topological domains in human LCLs. My analysis shows that elincRNAs are associated with high density of chromosomal contacts within TADs and are significantly enriched in loop anchors where promoter-enhancer interactions occur. I also find that they are strongly enriched in cohesin binding, supporting the idea that they may contribute to gene regulation by establishing contacts between regulatory elements and modulating chromosomal organization.

Results

Figures, tables and legends

Discussion

Materials and methods

Bibliography

[Darrow and Chadwick, 2013] Darrow, E. M. and Chadwick, B. P. (2013). Boosting transcription by transcription: enhancer-associated transcripts. *Chromosome Research*, 21(6):713–724.

[ENCODE Project et al., 2012] ENCODE Project, Bernstein, B. E., Birney, E., Dunham, I., Green, E. D., Gunter, C., and Snyder, M. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature*, 489(7414):57–74.

[Engreitz et al., 2016] Engreitz, J. M., Ollikainen, N., and Guttman, M. (2016). Long non-coding RNAs: spatial amplifiers that control nuclear structure and gene expression. *Nature Reviews Molecular Cell Biology*, 17(12):756–770.

[Gorkin et al., 2014] Gorkin, D. U., Leung, D., and Ren, B. (2014). The 3D genome in transcriptional regulation and pluripotency. *Cell Stem Cell*, 14(6):771–775.

[Guil and Esteller, 2012] Guil, S. and Esteller, M. (2012). Cis-acting noncoding RNAs: friends and foes. *Nature Structural & Molecular Biology*, 19(11):1068–1075.

[Hadjur et al., 2009] Hadjur, S., Williams, L. M., Ryan, N. K., Cobb, B. S., Sexton, T., Fraser, P., Fisher, A. G., and Merkenschlager, M. (2009). Cohesins form chromosomal cis-interactions at the developmentally regulated IFNG locus. *Nature*, 460(7253):410–413.

[Iyer et al., 2015] Iyer, M. K., Niknafs, Y. S., Malik, R., Singhal, U., Sahu, A., Hosono, Y., Barrette, T. R., Presner, J. R., Evans, J. R., Zhao, S., Poliakov, A., Cao, X., Dhanasekaran, S. M., Wu, Y.-M., Robinson, D. R., Beer, D. G., Feng, F., Iyer, H. K., and Chinnnaiyan, A. M. (2015). The Landscape of Long Noncoding RNAs in the Human Transcriptome. *Nat Genet.*, 47(3):199–208.

[Ji et al., 2016] Ji, X., Dadon, D. B., Powell, B. E., Fan, Z. P., Borges-Rivera, D., Shachar, S., Weintraub, A. S., Hnisz, D., Pegoraro, G., Lee, T. I., Misteli, T., Jaenisch, R., and Young, R. A. (2016). 3D Chromosome Regulatory Landscape of Human Pluripotent Cells. *Cell Stem Cell*, 18(2):262–275.

[Kornienko et al., 2013] Kornienko, A. E., Guenzl, P. M., Barlow, D. P., and Pauler, F. M. (2013). Gene regulation by the act of long non-coding RNA transcription. *BMC biology*, 11(1):59.

[Lupiáñez et al., 2016] Lupiáñez, D. G., Spielmann, M., and Mundlos, S. (2016). Breaking TADs: How Alterations of Chromatin Domains Result in Disease. *Trends in Genetics*, 32(4):225–237.

[Marques et al., 2013] Marques, A. C., Hughes, J., Graham, B., Kowalczyk, M. S., Higgs, D. R., and Ponting, C. P. (2013). Chromatin signatures at transcriptional start sites separate two

equally populated yet distinct classes of intergenic long noncoding RNAs. *Genome Biology*, 14(11):R131.

[Ong and Corces, 2014] Ong, C.-t. and Corces, V. G. (2014). CTCF : an architectural protein bridging genome topology and function. *Nature Publishing Group*, 15(4):234–246.

[Passarge, 1979] Passarge, E. (1979). Emil Heitz and the concept of heterochromatin: longitudinal chromosome differentiation was recognized fifty years ago. *American journal of human genetics*, 31(2):106–15.

[Pope et al., 2014] Pope, B. D., Ryba, T., Dileep, V., Yue, F., Wu, W., Denas, O., Vera, D. L., Wang, Y., Hansen, R. S., Canfield, T. K., Thurman, R. E., Cheng, Y., Gülsoy, G., Dennis, J. H., Snyder, M. P., Stamatoyannopoulos, J. A., Taylor, J., Hardison, R. C., Kahveci, T., Ren, B., and Gilbert, D. M. (2014). Topologically associating domains are stable units of replication-timing regulation. *Nature*, 515(7527):402–405.

[Rao et al., 2014] Rao, S. S. P., Huntley, M. H., Durand, N. C., Stamenova, E. K., Bochkov, I. D., Robinson, J. T., Sanborn, A. L., Machol, I., Omer, A. D., Lander, E. S., and Aiden, E. L. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*, 159(7):1665–1680.

[Rinn and Chang, 2012] Rinn, J. L. and Chang, H. Y. (2012). Genome regulation by long noncoding RNAs. *Annual review of biochemistry*, 81:145–166.

[Tamaru, 2010] Tamaru, H. (2010). Confining euchromatin/heterochromatin territory: Jumonji crosses the line. *Genes and Development*, 24(14):1465–1478.

[Yin et al., 2015] Yin, Y., Yan, P., Lu, J., Song, G., Zhu, Y., Li, Z., Zhao, Y., Shen, B., Huang, X., Zhu, H., Orkin, S. H., and Shen, X. (2015). Opposing roles for the lncRNA haunt and its genomic locus in regulating HOXA gene activation during embryonic stem cell differentiation. *Cell Stem Cell*, 16(5):504–516.